Incentive compatible reward schemes for labour-managed firms

Salvador Barberà[1], **Bhaskar Dutta**[2]

[1] Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
(e-mail: salvador.barbera@uab.es)
[2] Indian Statistical Institute, 7 SJS Sansanwal Maarg, New Delhi 110016, India
(e-mail: dutta@isid.ac.in)

Abstract. We consider a simple case of team production, where a set of workers have to contribute a single input (say labour) and then share the joint output amongst themselves. Different incentive issues arise when the skills as well as the levels of effort expended by workers are not publicly observable. We study one of these issues in terms of a very simple model in which two types of workers, skilled and unskilled, supply effort inelastically. Thus, we assume away the problem of moral hazard in order to focus on that of adverse selection. We also consider a hierarchical structure of production in which the workers need to be organised in two tiers. We look for reward schemes which specify higher payments to workers who have been assigned to the top-level jobs when the principal detects no lies, distribute the entire output in all circumstances, and induce workers to revel their true abilities. We contemplate two scenarios. In the first one, each individual worker knows only her own type, while in the second scenario each worker also knows the abilities of all other workers. Our general conclusion is that the adverse selection problem can be solved in our context. However, the range of satisfactory reward schemes depends on the informational framework.

Key words: Incentives, adverse selection, strategy-proofness, reward schemes, labour-managed firms

JEL classification: D82, J54, D20

We are most grateful to an anonymous referee, J. Crémer, M. Jackson, I. Macho, D. Pérez-Castrillo, and specially A. Postlewaite for very helpful discussions and suggestions.

1 Introduction

In the simplest cases of team production, there is a set of workers who each have to contribute a single input (say labour) and then share the joint output amongst themselves. Different incentive issues arise when the skills as well as the levels of effort expended by workers are not publicly observable. The issue of *moral hazard*, which appears whenever the supply of the input involves some cost, is well recognised in the literature.[1] In contrast, the problem of *adverse selection* which is caused by the presence of workers of differential abilities, seems to have been relatively neglected. The purpose of this paper is to study the possibility of designing suitable incentive schemes which will induce workers to reveal their true abilities.

We study this problem in terms of a very simple model in which *two* types of workers, skilled and unskilled, supply effort *inelastically*.[2] Thus, we assume away the problem of moral hazard in order to focus on the issues raised by adverse selection. We also consider a hierarchical structure of production in which the workers need to be organised in two tiers. The first-best outcome requires that only skilled workers be assigned to the *top* level jobs since these require special skills. Indeed, we specify that unskilled workers are more productive at the low level jobs. The adverse selection problem arises because skilled workers need to be paid more than unskilled workers when the principal[3] can verify that all workers have told the truth.

Since types are not observable, there is a need to design a system of payments which will induce workers to reveal their types correctly. Since the principal can observe the realized output, the payment schedule can be made contingent on realized output as well as on the assignment of tasks. A trivial way to solve the adverse selection problem is to distribute the realized output equally under all circumstances. It will then be in the interests of all workers to maximise total product, and hence to volunteer the true information about abilities so as to achieve an optimal assignment of tasks. However, this extreme egalitarianism may be inappropriate. For example, skilled workers may have better outside options and hence higher reservation prices than the unskilled workers.

Another trivial way to solve the adverse selection problem is to levy very harsh punishment on *all* workers whenever lies are detected. Observe that since the principal observes the realized output, she can detect lies whenever unskilled workers claiming to be skilled have been assigned to the top level jobs. However, such punishments imply that some output has to be destroyed. This will typically not be *renegotiationproof*. Therefore, we look for reward schemes which

[1] See for instance Sen (1966), Israelson (1980) or Thomson (1982) for related work on labour-managed firms. Groves (1973) and Holmstrom (1982) are a couple of papers which deal with the more general framework of teams.

[2] In the last section, we describe a more general model containing more than 2 types in which almost all our results remain valid.

[3] Notice that there is no actual principal as in the standard principal-agent models. Following standard practice in implementation theory, we use the term "principal" to represent the set of agreements or rules used by the workers to run the cooperative.

specify higher payments to workers who have been assigned to the top-level jobs when the principal detects no lies, and which distribute the entire output in all circumstances.

Our general conclusion is that the adverse selection problem can be solved in our context. However, the range of possible reward schemes depends on the informational framework. We contemplate two scenarios. In the first one, where each individual worker knows only her own type, there exist strategyproof (in fact even group strategyproof) reward schemes. But these schemes can only accomodate limited pay differentials between workers of different types. As we shall see, this implies the incompatibility of strategyproofness with some reasonable distributional principles. In the second scenario, each worker also knows the abilities of all other workers.[4] In this case, the class of reward schemes solving the adverse selection problem is much wider.

2 The formal framework

Let N be the set of n members of a cooperative enterprise. We assume that workers are of *two* types - *skilled* (or more able) and *unskilled* (or less able). T_1 will denote the set of skilled workers, who will also be called the Type 1 workers. T_2 will denote the set of unskilled workers, who will be labelled Type 2 workers. We assume that both sets are *nonempty* since an adverse selection cannot arise if one of the sets is empty. Note that the type of each worker is *private information*- there are no external characteristics which can be used to identify workers' types.

Two kinds of jobs need to be performed in order to produce output. One type of job is essentially a routine or mechanical activity, and does not require any special skills. So, both types of workers are equally proficient at performing this job, which will henceforth be labelled as J_2 or Type 2 job. In contrast, the *Type 1* job, to be denoted J_1, involves "managerial" responsibilities requiring some skill. Hence, these should ideally be performed by the Type 1 workers. However, if Type 2 workers are assigned to J_1, then they perform their job *inefficiently*, and are responsible for some loss of output. We model this by stipulating that output increases strictly when a Type 2 worker is shifted from the Type 1 job to the Type 2 job. We also assume that the *maximum* cardinality of J_1 is given by some number K, where $K \leq n$.[5] However, it turns out that except in Sect. 4, the possible restriction on the number of Type 1 positions does not affect any of our results.

Let t_{ij} denote the number of workers of type $i (i = 1, 2)$ employed in job $j (j = 1, 2)$. Hence, the "organizational structure" of the enterprise can be described by a vector $t = (t_{11}, t_{12}, t_{21}, t_{22})$. Let T denote the set of such vectors t with (i) $t_{11} + t_{21} \leq K$, and (ii) $t_{11} + t_{12} + t_{21} + t_{22} = n$. So, T represents the set of *feasible*

[4] Notice that an adverse selection problem arises even in this case since the information about other workers' types is not verifiable.

[5] Given our interpretation of jobs, this seems a natural restriction.

structures, with (i) expressing the requirement that no more than K workers can be in J_1, while (ii) states that all the n workers have to be employed.

We also assume that all workers supply one unit of effort *inelastically*. We are therefore assuming away the problem of *moral hazard*. We do this in order to focus on some of the issues raised by *adverse selection*.

Let $f(t)$ represent the function describing output produced by any particular structure. The following assumptions are made on the production function f.

Assumption 1: For all $t, t' \in T$,

(i) $f(t) = f(t')$ if $t_{11} = t'_{11}$ and $t_{21} = t'_{21}$.
(ii) $f(t) > f(t')$ if $t_{11} > t'_{11}$ and $t_{21} = t'_{21}$.
(iii) $f(t) > f(t')$ if $t_{11} = t'_{11}$ and $t_{21} < t'_{21}$.

Condition (i) in the Assumption says that if two structures differ only in the composition of workers performing Type 2 jobs, then the output produced must be the same. This expresses the notion that both skilled and unskilled workers are equally adept at performing the Type 2 job. Condition (ii) essentially captures the idea that skilled workers are more productive doing Type 1 jobs than Type 2 jobs provided no more than K workers are employed at Type 1 jobs. Conversely, Condition (iii) states that the unskilled workers are unsuitable for Type 1 jobs.

Notice that given Assumption 1, the total output produced by the enterprise is determined completely by the composition of workers performing Type 1 jobs. We will sometimes find it convenient to represent the output of the enterprise by $f(k, l)$, where k and l are respectively the numbers of workers in T_1 and T_2 doing Type 1 jobs.

An interesting special case of the general model, which will be used in the next section, is described below. Choose a vector $p = (p_1, p_2, p_3)$ with $p_1 > p_2 > p_3 \geq 0$, and a number $C > 0$. Then, in the p-model, the output produced is given by

$$f(k, l) = kp_1 + (n - k - l)p_2 + lp_3 - C \tag{1}$$

Equation (1) has the following interpretation. C represents the *fixed cost* of running the enterprise. Moreover, each worker in a Type 2 job has a productivity of p_2. In Type 1 jobs, the skilled workers have a productivity of p_1, while the unskilled workers have a productivity of p_3. Since $p_1 > p_2 > p_3$, it is easy to check that the p-model satisfies Assumption 1 above.

If workers' types were publicly observable, then upto K skilled workers would be assigned to Type 1 jobs, while the rest would be assigned to Type 2 jobs. However, since types are private information, the *principal* cannot adopt this naive procedure. So, she has to design a *reward scheme* or payment schedule which will induce workers to reveal their *true* types. Notice that since the principal can observe the organizational structure and the total output realized, the reward to each worker can be made contingent on output as well as the structure $t \in T$. In fact, the principal can, after observing output, actually infer the *number* of workers in T_2 who have actually lied and been assigned to J_1. Of course, the principal cannot infer *who* have lied. Nor can the prinicipal deduce anything about workers in T_1 who have falsely claimed to be in T_2 and hence been assigned to J_2. Nevertheless, it is apparent that the principal in this setting has more information than in the traditional implementation framework.

This suggests the following scenario. First, the principal announces the *assignment rule* which she will use to determine the production structure as a function of the information revealed by the individuals. Second, she also announces the reward scheme which make payments a function of (i) realized output (ii) the structure $t \in T$ which she will choose after hearing the vector of announcements by the workers.

Given the reward scheme, each worker announces his private information. As far as a worker's private information is concerned, we describe two alternative possibilities. In the first case, an individual only knows his or her *own* type. Naturally, in this case, an individual's announcement consists of a declaration of one's own type. The second case corresponds to that of complete information, where each individual knows every other worker's type. In the latter case, an announcement consists of a *profile* of types, one for each worker.

The announcements made by the workers together with the assignment rule chosen by the principal determines the organizational structure. The workers perform their assigned job, output is realized, and subsequently distributed according to the reward scheme announced by the principal. Notice that the organizational structure may be inoptimal if workers have lied about their types. For instance, if worker i *falsely* claims to be skilled, then he may be assigned to J_1, although he would be more productive in a Type 2 job.

The formal framework is as follows. The principal announces an *assignment rule* A which assigns each worker i to either J_1 or J_2 as a function of the information vector announced by the workers. She also announces a reward scheme, which is a pair of functions $\mathbf{r} = (r_1, r_2)$, where

$$r_i : \mathbb{R}_+ \times T \to \mathbb{R}_+. \tag{2}$$

Here, $r_1(y,t)$ is the reward to workers assigned to J_1, contingent on output being y, while $r_2(y,t)$ is the corresponding payment promised to workers assigned to Type 2 jobs. Remembering our earlier remark that output is completely specified by the composition of workers assigned to J_1, we will sometimes represent a reward scheme as $\{r_1(k,l), r_2(k,l)\}$, where k and l are the numbers of skilled and unskilled workers assigned to Type 1 jobs. This formulation assumes that the principal can infer how many unskilled workers have been assigned to Type 1 jobs. Note that knowledge of the production function is enough for this purpose.

Equation (2) also assumes that the principal has to employ *anonymous* schemes - the reward to workers i and j cannot differ if they are assigned to the same job. In particular, workers i and j may both have been assigned to J_2 even though i may have announced that she is *skilled* and j may have announced that she is *unskilled*.[6] In other words, agents' announcements about types matter only in so far as this influences the assignment to jobs. A more general approach[7] would have been to consider schemes in which worker i is paid more than worker j. Notice, however, that if workers announce only their own types, then the principal has no way of verfying whether i has announced the truth if she has been assigned to J_2. Hence, if i is paid more than j, then that would give j an incentive t o declare that she is skilled!

Of course, if worker j wrongly claims to be skilled, then she would also have to take into account the possibility that she is assigned to J_1. If she is indeed assigned to J_1, then the principal would detect that *someone* has lied, and then j (along with others assigned to J_1) would have to pay a penalty. The probability that j is assigned to J_1 depends on the number of other workers who have announced that they are skilled, the number of positions in J_1, and the tie-breaking ru le used by the principal. Clearly, non-anonymous schemes would have to satisfy very complicated schemes in order to be induce truthtelling as a dominant strategy. That is why we have chosen the simpler (but somewhat less general) approach of restricting attention to anonymous schemes.

We also consider the complete information case when workers announce entire type profiles. In this case, *other* workers' announcements could in principle be used to distinguish between two workers assigned to J_2. Here, non-anonymous schemes ca give rise to a differet problem. Suppose skilled worker i is assigned to J_2, and paid more than the unskilled workers. Then, the unskilled workers may have an incentive to declare i to be unskilled. This, by decreasing the amount paid to i will leave more to be distributed to the others. Notice again that there is no way in which the principal can verify that the others have told the truth about i.

In what follows, we will refer to an assignment rule and reward scheme as a *mechanism*.

Clearly, each specification of a mechanism gives rise to a normal form game in which the workers' strategies are to announce either their own types or an entire vector of types, depending upon the structure of information. We assume that the principal's primary objective is to choose mechanisms which will induce workers to reveal their private information truthfully in equilibrium. Of course, this involves the appropriate choice or specification of an equilibrium, depending upon the informational framework. In this paper, we focus on *strategyproof* mechanisms, that is mechanisms under which truthtelling is a *dominant strategy*, in the case when workers know only their own types. In the complete information framework, we restrict attention to Nash equilibria and undominated strategy equilibria. In other words, we are interested in the issue of designing mechanisms under which the sets of these equilibria will coincide with truthtelling or strategies which are equivalent to truthtelling.

---

[6] Notice that this issue matters only for workers assigned to J_2 since all workers assigned to J_1 must have announced that they are skilled.

[7] We are grateful to the anonymous referee for pointing out the need to clarify this issue.

While these concepts are defined rigorously in subsequent sections, we specify below some restrictions which will be imposed on all reward schemes. These restrictions essentially ensure that the problems we are studying are nontrivial.[8]

Definition 1. *A reward scheme* $\mathbf{r}$ *is admissible if*

(i) $(k+l)r_1(k,l)+(n-k-l)r_2(k,l)=f(k,l)\ \forall k,l$ *such that* $k+l \leq K$
(ii) $r_1(k,0) > r_2(k,0)\ \forall k \leq K$.

Remark 1. In this paper, we are going to restrict attention to admissible reward schemes. Henceforth, reward schemes are to be interpreted as admissible reward schemes.

Feasibility requires that the sum of the payments made to the workers never exceeds realized output. Condition (i) goes a step further, and insists that the principal can never destroy output. As we have mentioned earlier, a feasible reward scheme which leaves some surplus is open to renegotiation.

Condition (ii) states that if the principal observes a level of output which confirms that all workers assigned to J_1 are skilled, then these workers must be paid more than the rest. Notice that unless skilled workers are paid *at least as much* as unskilled workers, the former will not have any incentive to reveal their true types. It is also obvious that under the reward scheme which *always* distributes output equally amongst all workers, the adverse selection problem disappears. The imposition of Condition (ii) can be thought of as a search for "non-trivial" incentive compatible reward schemes. Also, such differentials may be necessary because of superior outside options for the skilled workers.

3 Strategyproof reward schemes

In this section, we first define the conditions of strategyproofness and group strategyproofness. We go on to derive a necessary and sufficient condition for strategyproof reward schemes. We then show that the class of such schemes is nonempty - indeed, we prove a stronger result by constructing a reward scheme which is *group strategyproof*. Finally, we explore the possibility of constructing strategyproof schemes which are also "nice" from an ethical point of view.

When workers only know their own types, an *announcement vector* $a = (a_1, \ldots, a_n)$ is an n-tuple of messages sent by the workers, each a_i representing worker i's claim about his type. We will use $a_i = 1$ to denote the claim that i is skilled, while $a_i = 2$ will denote the claim that i is unskilled. Given the assignment rule A employed by the principal, an announcement vector a generates a structure $t = A(a)$. The reward scheme $\mathbf{r}$ applied to t and the realized output then gives the payoff vector $R(a, \mathbf{r})$ associated with a. This is given by

[8] Also, notice that our formulation rules out the use of various ad hoc features such as *tailchasing* which are often incorporated in game forms employed in the traditional literature on implementation. For a review of the criticism against the use of these features, see Dutta(1997), Jackson(1992), Moore(1992).

$$R^i(a,\mathbf{r}) = \begin{cases} r_1(k(a), l(a)) & \text{if } i \text{ is assigned to Type 1 job} \\ r_2(y, A(a)) & \text{otherwise} \end{cases} \quad (3)$$

where $k(a), l(a)$ are the number of skilled and uskilled workers assigned to J_1 according to the anouncement a.[9] Notice that when workers announce only their own types, the principal has essentially no freedom in so far as the assignment rule is concerned. If some workers declare that they are skilled, the principal must treat these claims as if they are true since she cannot detect lies *before* the output is realized. Hence, the "best" chance of achieving efficiency is to assign upto K workers to J_1 from amongst those workers who claim to be in T_1.[10] So, the principal has to use only the *reward scheme* to induce workers to tell the truth. In view of this, we will define strategyproofness to be a property of reward schemes, although strictly speaking it is the combination of the assignment rule and the reward scheme which defines the appropriate game.

Let a^* denote the vector of *true* types of workers.

For any coalition S, a vector a will sometimes be denoted as (a_S, a_{-S}).

Definition 2. *For any coalition S, a_S is a coalitionally dominant strategy profile under reward scheme $\mathbf{r}$ iff*

$$\sum_{i\in S} R^i(a_S, a_{-S}, \mathbf{r}) \geq \sum_{i\in S} R^i(\hat{a}_S, a_{-S}, \mathbf{r})\ \forall \hat{a}_S, \forall a_{-S}.$$

So, a_S is a coalitionally dominant strategy profile for coalition S if it is a best reply to *any* vector of strategies chosen by workers outside the coalition. When the coalition S consists of a single individual, we will use the terminology *dominant strategy*.

Definition 3. *A reward scheme $\mathbf{r}$ is group-strategyproof if for all coalitions S, a_S^* is a coalitionally dominant strategy profile under $\mathbf{r}$.*

This definition assumes the possibility of side payments within any coalition. If side payments are not possible, then the corresponding definition of group strategyproofness would be *weaker*. Since our result on group strategyproofness (Proposition 2) demonstrates the existence of group strategyproof schemes, we use the definition which leads to a stronger concept.

Definition 4. *A reward scheme $\mathbf{r}$ is strategyproof if for all individuals i, a_i^* is a dominant strategy under $\mathbf{r}$.*

The following notation will be used repeatedly. Call a pair of integers (k,l) *permissible* if $k + l \leq K$ and $k \geq 1, l \geq 1$.

---

[9] Whenever there is no confusion about the anouncement vector a, we will simply write $r_i(k,l)$ instead of $r_i(k(a), l(a))$.

[10] If more than K workers claim to be in T_1, then the principal has to use some rule to select a set of K workers. We omit any discussion of these selection rules since the results of this section are not affected by the choice of the selection rule.

Proposition 1. *An admissible reward scheme* **r** *is strategyproof iff* **r** *satisfies the following conditions for all permissible pairs* (k, l)*:*

$$r_2(k-1, l) \leq r_1(k, l) \leq r_2(k, l-1) \tag{4}$$

Proof. Consider any **r**, and suppose for some permissible pair (k, l), $r_2(k-1, l) > r_1(k, l)$. Consider a^* such that $|T_1| = k$, and let $i \in T_1$. Consider a such that $|\{j \in T_2 | a_j = 1\}| = l$ and $a_m = a_m^* \; \forall m \in T_1$. That is, all skilled workers declare the truth about their types, but exactly l unskilled workers claim to be skilled. Then, $R^i(a, \mathbf{r}) = r_1(k, l)$. Suppose i deviates and announces $\bar{a}_i = 2$. Then, $R^i(\bar{a}_i, a_{-i}, \mathbf{r}) = r_2(k-1, l) > R^i(a, \mathbf{r})$. But, then **r** is not strategyproof.

Suppose now that $r_1(k, l) > r_2(k, l-1)$. Let a^* be such that T_1 contains k workers. Consider a such that $(l-1)$ unskilled workers declare themselves to be skilled, all other workers telling the truth. Let $j \in T_2, a_j = a_j^*$. Then $R^j(a_j^*, a_{-j}, \mathbf{r}) = r_2(k, l-1) < R^j(\bar{a}_j, a_{-j}, \mathbf{r}) = r_1(k, l)$ when $\bar{a}_j = 1$. Then, **r** is not strategyproof. These establish the necessity of (4).

We now want to show that if **r** satisfies (4), then it is strategyproof.

Suppose **r** satisfies (4). If for some i, a_i^* is *not* a dominant strategy, then there are two possible cases.

Case(i): $i \in T_1$. Let $\bar{a}_i = 2$. Then, there is a_{-i} such that

$$R^i(\bar{a}_i, a_{-i}, \mathbf{r}) > R^i(a_i^*, a_{-i}, \mathbf{r}) \tag{5}$$

But, (5) is not possible if $r_2(k-1, l) \leq r_1(k, l)$ for each permissible pair (k, l).

Case (ii): $i \in T_2$. Let $\bar{a}_i = 1$. Suppose there is a_{-i} such that

$$R^i(\bar{a}_i, a_{-i}, \mathbf{r}) > R^i(a_i^*, a_{-i}, \mathbf{r}) \tag{6}$$

But, (6) is not possible in view of $r_1(k, l) \leq r_2(k, l-1)$ from (4). So, a_i^* must be a dominant strategy for all i. □

In the next Proposition, we construct a group strategyproof reward scheme. The reward scheme has the following features. The payment made to an individual in J_1 exceeds the payment made to an individual in J_2 by a "small" amount when no lies are detected. If the principal detects any lie, then the output is distributed equally. The proof essentially consists in showing that provided the difference in payments to individuals in J_1 and J_2 are small enough, no group can gain by misrepresenting their types.

Proposition 2. *There exists a group-strategyproof reward scheme.*

Proof. Let f be the production function. Define the following:

$$\begin{aligned} a(k,l) &= \frac{f(k,l)}{n} \ \forall k,l \text{ such that } k+l \leq K \\ \gamma &= \min_k \{k(k+1)[a(k+1,0) - a(k,0)]\} \\ \epsilon &= \min_k \{f(k,0) - f(k,1)\} \\ \delta &= \frac{1}{n} \min(\epsilon, \gamma) \end{aligned}$$

Consider the following reward scheme $\mathbf{r}$.

$$\begin{aligned} r_1(k,0) &= a(k,0) + \frac{n-k}{k}\delta \\ r_2(k,0) &= a(k,0) - \delta \\ \forall i = 1,2, r_i(k,l) &= a(k,l) \ \forall \text{ permissible pairs } k,l \text{ such that } l \geq 1 \end{aligned}$$

Claim 1. $r_1(k,l)$ is monotonically increasing in k.

The claim is obviously true for all $l \geq 1$ since $f(k,l)$ is increasing in k, and since $r_1(k,l) = a(k,l)$. So, it is sufficient to prove that $r_1(k+1,0) \geq r_1(k,0) \ \forall k < K$. To see this, note that

$$\begin{aligned} r_1(k+1,0) - r_1(k,0) &= a(k+1,0) - a(k,0) - \frac{\delta n}{k(k+1)} \\ &\geq 0 \text{ since } n\delta \leq \gamma. \end{aligned}$$

Claim 2. $\mathbf{r}$ is group-strategyproof.

Take any coalition S. We need to show that no matter what announcements are made by $(N \setminus S)$, a_S^* is a best reply of S.

Suppose not. Then, there is a_S, a_{-S} such that

$$\sum_{i \in S} R^i(a_S, a_{-S}, \mathbf{r}) > \sum_{i \in S} R^i(a_S^*, a_{-S}, \mathbf{r}) \tag{7}$$

This cannot hold if there is $i \notin S$ such that $i \in T_2 \cap J_1$. For, then the "average rule" applies, and any deviation from the truth by S can only reduce aggregate output, and hence their own share.

So, without loss of generality, let $a_{-S} = a_{-S}^*$. First, suppose there is $i \in S$ such that $a_i^* = 2$, but $a_i = 1$. Then, a lie is detected, and the average rule is applied. However, the choice of δ guarantees that $r_2(k,0) \geq a(k,1) \geq a(k',l)$ $\forall l \geq 1, \forall k' \leq k$. Since $r_1(k,0) > r_2(k,0)$, *no* individual in S can be better-off.

So, the only remaining case is when $\forall i \in S$, $a_i \neq a_i^*$ implies $a_i^* = 1$ and $a_i = 2$. However, given Claim 1, $r_1(k,0) \geq r_1(k',0) \ \forall k' \leq k$. Also, $r_2(k,0) \geq r_2(k',0)$. So, again this deviation from a_S^* cannot benefit anyone in S.

So, $\mathbf{r}$ is group-strategyproof. □

Since strategyproof reward schemes exist, a natural question to ask is whether it is possible to construct such schemes which are also satisfactory from other perspectives. This is what we pursue in the rest of this section.

First, one ethical principle which is appealing in this context is that workers whose contributions to production are proven to be in accordance with their declared types should not be punished for any loss of output. That is, consider $f(k,0)$ and $f(k,l)$. Although $f(k,0) > f(k,l)$, workers who have been assigned to Type 2 jobs are not responsible for the loss of output. Hence, they should not be punished. We incorporate this principle in the following Axiom.

Axiom 1. $r_2(k,0) \leq r_2(k,l)$ for all permissible pairs k, l.

Unfortunately, it is not possible to construct strategyproof reward schemes which always satisfy Axiom 1. This is the content of the next proposition.

Proposition 3. *There exist production functions such that no strategyproof reward scheme satisfies Axiom 1.*

Proof. Consider the p-model defined in the previous section with $p_3 = 0$. To simplify notation, also assume that $C = 0$.

Let $\mathbf{r}$ be a strategyproof scheme satisfying Axiom 1. Denote $r_2(1,0) = \mu$.

Since $\mathbf{r}$ is strategyproof, we must have $\mu \geq r_1(1,1) \geq r_2(0,1)$. From Axiom 1, $r_2(0,1) \geq r_2(0,0)$. Since $r_2(0,0) = p_2$, we must have

$$\mu \geq p_2 \tag{8}$$

Choose any $i \leq K - 1$. Then,

$$\begin{aligned} (i+1)r_1(1,i) + (n-i-1)r_2(1,i) &= p_1 + (n-i-1)p_2 \\ \text{or } (1+i)r_1(1,i) &= p_1 - (n-i-1)[r_2(1,i) - p_2] \end{aligned}$$

Also, $r_2(1,i) \geq \mu \geq p_2$ from Axiom 1 and (8). Hence,

$$(1+i)r_1(1,i) \leq p_1 - (n-1-i)(\mu - p_2) \tag{9}$$

Since $\mathbf{r}$ is strategyproof, $r_1(1,i) \geq r_2(0,i)$. Also, from Axiom 1, $r_2(0,i) \geq r_2(0,0) = p_2$. Using $r_1(1,i) \geq p_2$ and (9), we get

$$p_1 - (n-1-i)(\mu - p_2) \geq (1+i)p_2 \tag{10}$$

Since $\mu \geq p_2$, this yields

$$p_1 \geq (1+i)p_2 \tag{11}$$

Obviously, a p-model can be specified for which this is not true.

This shows that strategyproofness and Axiom 1 are not always compatible. □

Axiom 1 imposed a restriction on the nature of possible punishments incorporated in reward schemes. Another restriction which one may want to impose on reward schemes is the principle of workers being paid "according to contribution" when the principal detects no lies. Of course, this principle is not always enforceable for the simple reason that the production function may be such that workers' marginal contributions do not add up to the gross output. However, one case in which this principle is a priori feasible is when the production function is described by the p-model. Here, the principle of "payment according to contribution" takes a simple form. For each value of $k \leq K$, one should have $r_1(k,0) = p_1 - \frac{C}{n}$ and $r_2(k,0) = p_2 - \frac{C}{n}$. In other words, all workers are paid their marginal product minus an equal share of the fixed cost. Unfortunately, we show below that the requirement of strategyproofness is not always compatible with this principle of payment.

Proposition 4. *There exists a p-model and a size of society such that the principle of "payment acording to contribution" is not strategyproof.*

Proof. Define for $i = 1, 2, 3$, $\bar{p}_i = p_i - \frac{C}{n}$. Clearly, $\bar{p}_1 > \bar{p}_2$.

Suppose $\mathbf{r}$ is strategyproof and satisfies the principle of payment according to contribution. So, for all $k \leq K$ and $i = 1, 2$, we must have $r_i(k,0) = \bar{p}_i$. From (4), $r_1(k,1) \leq r_2(k,0) = \bar{p}_2$. Since $(k+1)r_1(k,1)+(n-k-1)r_2(k,1) = k\bar{p}_1+\bar{p}_3+(n-k-1)\bar{p}_2$, we have $r_2(k,1) = \bar{p}_2+\frac{\Delta(k)}{n-k-1}$, where $\Delta(k) = k(\bar{p}_1 - r_1(k,1))+\bar{p}_3 - r_1(k,1)$. Since $(\bar{p}_1 - r_1(k,1)) > 0$, there exists a value of k, say k^*, such that $\Delta(k^*) > 0$. Hence, $r_2(k^*,1) > \bar{p}_2$.

But this contradicts the requirement that

$$r_2(k^*,1) \leq r_1(k^*+1,1) \leq r_2(k^*+1,0) = \bar{p}_2 . \qquad \square$$

4 The complete information framework

In the last section, we showed that there are non-trivial strategyproof schemes. Unfortunately, Propositions 3 and 4 show that such schemes may fail to satisfy additional attractive properties. This provides us with the motivation to examine whether an incentive requirement weaker than strategyproofness widens the class of permissible schemes. This is the avenue we pursue here by examining the scope of constructing reward schemes which induce workers to reveal their true information as equilibria in games of complete information.[11]

When each worker knows other workers' types, the principal can ask each worker to report a *type profile*, although of course she may not always utilise all the information. Let $a^i = (a^i_1, \ldots, a^i_n)$ be a typical report of worker i, with $a^i_j = 1$ denoting that i declares j to be in T_1. Similarly, $a^i_j = 2$ represents the statement that i declares j to be in T_2. Let $\mathbf{a} = (a^1, \ldots, a^n)$ denote a typical vector of announced type profiles. Let $m = (A, \mathbf{r})$ be any mechanism where A is the assignment rule specifying whether worker i is in J_1 or J_2 given workers' announcements $\mathbf{a}$. Letting $A(\mathbf{a})$ denote the structure produced when workers an-

[11] Actually, we are interested in a stronger requirement. In line with traditional implementation theory, we also want to ensure that truthtelling and strategies equivalent to truthtelling are the *only* equilibria.

nounce **a** and the principal uses the mechanism **m**, the payoff function of the corresponding game is given by[12]

$$R^i(a, \mathbf{m}) = \begin{cases} r_1(k(\mathbf{a}), \mathbf{l}(\mathbf{a})) & \text{if } i \text{ is assigned to} J_1 \\ r_2(k(\mathbf{a}), \mathbf{l}(\mathbf{a})) & \text{otherwise} \end{cases} \tag{12}$$

where $k(\mathbf{a}), \mathbf{l}(\mathbf{a})$ are the number of skilled ad uskilled workers assigned to J_1 respectively corresponding to the anouncementy vector **a**.[13]

Definition 5. *Given a mechanism* **m**, *an announcement* a^i *is undominated for worker i if there is no announcement* $\bar{a}^i$ *such that for all* a^{-i}, $R^i((\bar{a}^i, a^{-i}), \mathbf{m}) \geq R^i((a^i, a^{-i}), \mathbf{m})$ *with strict inequality for some* $\tilde{a}^{-i}$.

Definition 6. *Given a mechanism* **m**, *two announcement vectors* **a** *and* $\hat{\mathbf{a}}$ *are equivalent if* $R^i(\mathbf{a}, \mathbf{m}) = R^i(\hat{\mathbf{a}}, \mathbf{m})$ *for all* i.

Notice that *all* announcement vectors will be equivalent if the principal uses an assignment rule which is completely insensitive to workers' announcements. Hence, in order to ensure a satisfactory or non-trivial solution to the incentive problem, we need to ensure that only "sensible" assignment rules are used. This provides the motivation for the following definition.

Definition 7. *An assignment rule is seemingly efficient if corresponding to any announcement vector* **a** *satisfying* $a^i = a^j$ *for all* $i, j \in N$, *up to* K *workers declared to be in* T_1 *by all workers are assigned to* J_1 *and all the rest are assigned to* J_2.

The principal of course has no way of verifying whether workers have told the truth or not until the output has actually been realized. However, if all workers unanimously announce the same type profile, then the principal has no basis for disbelieving this announcement. The assignment in this case should assign only workers declared to be in T_1 to J_1. Of course, at most K such workers can be assigned to J_1. Notice that the definition places no restriction on how assignments are made when workers do not make the same anouncement. So, it is a very weak restriction.

In this section, we are interested in the Nash equilibria and undominated strategy equilibria[14] of mechanisms which use seemingly efficient assignment rules. Let $NE(\mathbf{m})$ and $UD(\mathbf{m})$ denote the set of Nash equilibria and undominated strategy equilibria of the mechanism **m**.

Definition 8. *A reward scheme* **r** *is implemented in Nash equilibrium (respectively undominated strategies) with a seemingly efficient assignment rule A if there is a mechanism* **m** *such that for* $\mathbf{m} = (A, \mathbf{r})$, $NE(\mathbf{m})$ *(respectively* $UD(\mathbf{m})$*) consist of truthtelling and strategies which are equivalent to truthtelling.*

[12] Note that in contrast to the incomplete information framework, the principal does have some freedom about the assignment rule. That is why we have explicitly introduced the mechanism **m** in the notation.

[13] To simplify notation, we will omit the dependence of k, l on the announcement vector **a**.

[14] An undominated strategy equilibrium is one in which no worker is using a dominated strategy.

Let $\mathbf{r}$ be implemented in Nash equilibrium with a seemingly efficient assignment rule according to the definition given above. Then, at any equilibrium announcement, the "correct" or "efficient" assignment will be made. Furthermore, workers in J_1 will be paid $r_1(k,0)$ while workers in J_2 will be paid $r_2(k,0)$ where $|J_1| = k$. An exactly similar interpretation is valid if $\mathbf{r}$ is implemented in undominated strategies. Thus, if the class of implementable reward schemes is rich enough, then the principal can ensure payments according to various desirable principles, apart from achieving the maximum possible output given workers' true types and the production function.

In our first proposition in this section, we identify sufficient conditions on the production function which ensure that a rich class of anonymous reward schemes are Nash implementable with a seemingly efficient rule. [15]

Proposition 5. *Suppose either (i)* $K < n$ *or (ii)* $\frac{f(k,n-k)}{n} < f(k-1, n-k)$ *for all* k. *Let* $\mathbf{r}$ *satisfy the following:*

(i) $r_1(k,0) > r_2(k-1,0)$ *for all* $k \leq K$
(ii) $r_1(k,l) = 0$ *and* $r_2(k,l) = \frac{f(k,l)}{n-k-l}$ $\forall l \geq 1$.

Then, $\mathbf{r}$ *is implementable in Nash equilibrium with seemingly efficient assignment rule.*

Proof. Let $\mathbf{r}$ be any reward scheme satisfying (i) and (ii). Consider the following assignment rule A. For all $\mathbf{a}$, let $T_1(\mathbf{a}) = \{i \in N | a_i^i = 1\}$. Without loss of generality, let $T_1(\mathbf{a}) = \{1, 2, \ldots, L\}$. If $L \leq K$, then all $i \in T_1(\mathbf{a})$ are assigned to J_1. If $L > K$, then $\{1, 2, \ldots, K\}$ are assigned to J_1. So, the assignment rule only depends on what each individual reports about herself. If no more than K workers claim to be in T_1, then they are all assigned to J_1. If more than K workers claim to be skilled, then the first K workers are assigned to J_1.

It is easy to check that this assignment rule is *seemingly efficient*.

Let $a^* = (a_1^*, \ldots, a_n^*)$ be the vector of true types. We first show that any $\mathbf{a}$ such that $a_i^i = a_i^*$ is a Nash equilibrium.

Suppose $i \in T_1$. Then, either (i) i is assigned to J_1 or (ii) $T_1(\mathbf{a})$ contains more than K workers and i is assigned to J_2. Now consider any deviation $\hat{a}^i$ such that $\hat{a}_i^i \neq a_i^*$. If (i) holds, then i' payoff is $r_1(k,0)$ before the deviation, and either $r_2(k,0)$[16] or $r_2(k-1,0)$ after the deviation. In either case, i's deviation is not profitable. If (ii) holds, then i's deviation does not change the outcome.

Suppose now that $i \in T_2$. Then, i's payoff when all workers tell the truth is $r_2(k,0)$. Consider any deviation $\hat{a}^i$ such that $\hat{a}_i^i = 1$. Either this does not change the assignment (if i is not amongst the first K workers who declare they are in T_1) or i is assigned to J_1. But, then since $r_1(k,1) = 0$ $\forall k$, i will not deviate.

Now, we show that any $\mathbf{a} \in NE(\mathbf{m})$ must produce the same payoff vector as the truth.

[15] We are most grateful to A.Postlewaite for suggesting the mechanism used in the proof of the proposition.

[16] i's payoff could be $r_2(k,0)$ if more than K workers had originally declared themselves to be skilled. Of course, in this case $k = K$.

Assume first that $K < n$. Let $\mathbf{a} \in NE(\mathbf{m})$, and suppose that there is $i \in T_2$ such that $a_i^i = 1$. If i is not assigned to J_1, then i's announcement of a_i^i instead of the truth does not change the outcome. If i is not assigned to J_1, then $R^i(\mathbf{a}, \mathbf{m}) = r_1(k, l) = 0$. But, i can deviate by announcing $\hat{a}_i^i = 2$. Then, i's payoff would be strictly positive.

So, if $\mathbf{a} \in NE(\mathbf{m})$, then T_2 must be assigned to J_2. Consider now $i \in T_1$, and suppose $a_i^i = 2$. If i deviates and announces $\hat{a}_i^i = 1$, then either (i) i is assigned to J_1 or (ii) i is not amongst the first K workers in T_1. If i is not assigned to J_1 after the deviation, then she must be better off, so that in case (i), $\mathbf{a}$ cannot be a Nash equilibrium. In case (ii), $\mathbf{a}$ gives the same outcome as the truth.

So, this shows that when $K < n$, any $\mathbf{a} \in NE(\mathbf{m})$ is equivalent to the truth. Suppose now that $K = n$, but $\frac{f(k,n-k)}{n} < f(k-1, n-k)$ for all k.

The only remaining case we have to consider is if $a^i = 1$ for all $i \in N$. Then, for all $i \in N$, $R^i(\mathbf{a}, \mathbf{m}) = \frac{f(k,n-k)}{n}$ for some k. But, then some $i \in T_1$ can deviate and announce $\hat{a}^i = 2$. Then, i's payoff will be $f(k-1, n-k)$. This is a profitable deviation for i. □

Remark 2. An anonymous referee has pointed out that the reward schemes incorporate very heavy punishment since $r_1(k, l) = 0$ for all $l \geq 1$. However, note that this provision will apply only out of equilibrium. Thus, the only stipulation on the reward scheme applying to equilibrium messages is that $r_1(k, 0) > r_2(k-1, 0)$ for all $k \leq K$. Since this is a very weak requirement, Proposition 5 shows that the planner can implement a large class of anonymous reward schemes.

Notice that the smaller is n, the more restrictive is the condition that $\frac{f(k,n-k)}{n} < f(k-1, n-k)$. In our next proposition, we show that practically all reward schemes can be implemented in undominated strategies without this restriction on the production function, provided $K = n$.

Proposition 6. *Let $K = n$. Let $\mathbf{r}$ satisfy the following.*

(i) $r_1(k, 0)$ and $r_2(k, 0)$ are increasing in k.
(ii) $r_1(k, l) = r_2(k, l) = \frac{f(k,l)}{n}$ for all $l \geq 1$.

Then, $\mathbf{r}$ is implementable in undominated strategies.

Proof. Consider the following assignment rule A. For any $\mathbf{a}$, define $U_1(\mathbf{a}) = \{j \in N \mid a_j^i = 1\ \forall i \in N\}$. So, the set $U_1(\mathbf{a})$ is the set of workers who are unanimously declared to be in T_1. Then, $A(\mathbf{a})$ assigns all workers in $U_1(\mathbf{a})$ to J_1, all other workers being assigned to J_2.

Let a^* be the vector of true types.

Step 1. Let $i \in T_1$. Then, the only undominated strategy of i is to announce a^*.

To see this, suppose $a^i \neq a^*$. There are two possible cases. Either (i) there is j such that $a_j^* = 1$ and $a_j^i = 2$ or (ii) there is j such that $a_j^* = 2$ and $a_j^i = 1$.

In all cases, we need only consider announcement vectors in which all other workers have declared j to be in T_1. Otherwise, i cannot unilaterally change j's assignment.

In case (i), consider first $j = i$, that is i lies about herself. Then, i is assigned to J_2. If some unskilled worker is assigned to J_1, then the "average rule" applies. Then, i does strictly better by announcing the truth about oneself since this increases aggregate output and hence the average.

If no unskilled worker is assigned to J_1, then the same conclusion emerges from the fact that $r_1(k,0) > r_2(k,0) \geq r_2(k-1,0)$.

Suppose now that $j \neq i$. Then, i's deviation to the truth about j is strictly beneficial when some unskilled worker is assigned to J_1. For then the average rule applies and aggregate output increases when j is assigned to J_1. To complete this case, note that i never loses by declaring the truth about j since $r_1(k,0)$ is increasing in k.

Consider now Case (ii). Suppose some unskilled worker other than j is assigned to J_1. Then, i's truthful declaration about j increases aggregate output, and hence i's share through the average rule. If no unskilled worker other than j is assigned to J_1, then again i gains strictly since $r_1(k,0) > \frac{f(k,0)}{n} > \frac{f(k,1)}{n}$.

This completes the proof of Step 1.

Step 2. If $i \in T_2$, and if a^i is undominated, then $a^i_j = 1$ for all $j \in T_1$.

Suppose $a^i_j = 2$ for some $j \in T_1$. Again, we need only consider announcement vectors in which all other workers declare j to be in T_1. If some unskilled worker is assigned to J_1, then i gains by declaring the truth about j since $f(k,l) > f(k-1,l)$ and the average rule applies. If only skilled workers are assigned to J_1, then i cannot lose by telling the truth since $r_2(k,0)$ is increasing in k.

This completes the proof of Step 2.

From Steps 1 and 2, $U_1(\mathbf{a}) = T_1$ whenever workers use undominated strategies. Hence, all workers in T_1 will be assigned to J_1 ad all workers in T_2 will be assigned to J_2. □

Remark 3. Notice that while truthtelling is the *only* undominated strategy for individuals in T_1, individuals in T_2 may falsely declare an *unskilled* worker i to be skilled at an undominated strategy. However, this lie or deception does not matter since some $j \in T_1$ will reveal the truth about i. Hence, Proposition 6 shows that for a very rich class of anonymous reward schemes, the outcome when individuals use undominated strategies is equivalent to truthtelling. Of course, this remarkably permissive conclusion is obtained at the cost of a strong restriction on the class of production functions since the proposition assumes that $K = n$. If $K < n$, then workers in T_1 may have to "compete" for the positions in J_1. This implies that declaring another Type 1 worker to be in T_2 is no longer a dominated strategy for some worker in T_1.

5 Conclusion

In this paper, we have used a very simple model in which incentive issues raised by adverse selection can be discussed. The main features of the model are the presence of two types of workers as well as two types of jobs. We conclude by pointing out that our results do not really depend on there being *two* types of workers and jobs. The model can easily be extended to the case of k types of workers and jobs, provided an assumption analogous to Assumption 1 is made. What we need to assume is that workers of Type i are most productive in jobs of type i. They are as productive as workers of Type $(i+j)$ in jobs of type $(i+j)$, and *less* productive in type $(i-j)$ jobs than in type $(i+j)$ jobs. With this specification and the assumption that despite possible capacity restrictions on jobs of a particular type , the first best assignment never places a worker of type i in a job of type $(i-j)$, the principal can still detect whether workers of a particular type have claimed to be of a higher type. Notice that except in Proposition 1, the specification of the reward schemes did not need knowledge of *how many* workers had lied. It was sufficient for the principal to know that realized output was lower than the expected output. Hence, obvious modifications of the reward schemes and assignment rules will ensure that Propositions 2, 5 and 6 can be extended to the k type case. Of course, Propositions 3 and 4 are true since they are in the nature of counterexamples. It is only in the case of Proposition 1 that the reward scheme needs to use detailed information on the *number* of people who have lied. This came for free in the two-type framework, given assumption 1. In the general k type model, we would need to assume that the principal can on the basis of the realized output, "invert" the production function and find out how many workers of each type have lied and claimed to be of a higher type. Note that this will be generically true for the class of production functions satisfying the extension of Assumption 1 outlined above.

References

Dutta, B. (1997) Reasonable mechanisms and Nash implementation. In: Arrow, K.J., Sen, A.K., Suzumura, K. (eds.) *Social Choice Theory Reexamined*. Macmillan, London

Jackson, M. (1992) Implementation in undominated strategies: A look at bounded mechanisms. *Review of Economic Studies* 59: 757-75

Groves, T. (1973) Incentives in teams. *Econometrica* 41: 617-31

Holmstrom, B. (1982) Moral hazard in teams. *The Bell Journal of Economics* 13: 324-340

Moore, J. (1992) Implementation, contracts and renegotiation in environments with complete information. In: Laffont, J.J. (ed.) *Advances in Economic Theory*. Cambridge University Press, Cambridge

Israelson, D.L. (1980) Collectives, communes and incentives. *Journal of Comparative Economics* 4: 99-121

Thomson, W. (1982) Information and incentives in labour-managed economies. *Journal of Comparative Economics* 6: 248-268