Rational Sabotage in Cooperative Production*

Carmen Beviá

Departament d'Economia i d'Història Econòmica and CODE

Universitat Autònoma de Barcelona

Luis C. Corchón

Departamento de Economía. Universidad Carlos III de Madrid

October, 2002.

Abstract

In this paper we consider a model of cooperative production in which rational agents have the possibility to engage in sabotage activities that decrease output. It is shown that sabotage depends on the interplay between the degree of congestion, the technology of sabotage, the number of agents the degree of meritocracy and the form of the sharing rule. In particular it is shown that, ceteries paribus, meritocratic systems give more incentives to sabotage than egalitarian systems. We address two questions: The degree of meritocracy that is compatible with absence of sabotage and the existence of a Nash equilibrium with and without sabotage.

*We thank Paco Alcala, Clara Eugenia Garcia, Shasikanta Nandeibam, Ines Macho-Stadler, David Perez-Castrillo and Clara Ponsati for their helpful comments. The first author acknowledges finantial support from DGCYT and Direcció General de Recerca under projects BEC2002-02130, and 2000SGR-00054 and the second author from CAYCIT PB98-0024.

1. Introduction

Economics is the study of resource allocation under certain constraints. Historically, the first constraint considered was feasibility, i.e. the quantity of the good to be allocated can not exceed its available quantity. As economics developed, other constraints were considered, for instance Incentive Constraints: Any economic system must not give incentives to the agents to transmit wrong information -adverse selection- or to take socially unwanted actions -moral hazard.

In this paper we want to broad the scope of incentive theory by considering situations in which agents can sabotage production by destroying other people inputs. Thus, the input supplied by an agent reflects her effort and the sabotage done by others on this agent. We assume that sabotage is undetectable because effort and the technology of sabotage are not contractible.[1] On the contrary, inputs are contractible. Our first concern is to look for distribution rules that do not give incentives to rational agents to sabotage production.[2] An example of how sabotage may arise as a rational action is the following.

Two people are collecting grapes. Andy collects white grapes -whose quantity is denoted by R_1- and Beth collects red grapes, whose quantity is denoted by R_2. These grapes are transformed in wine -denoted by $Y-$ according to the production function $Y = (R_1 + R_2)^{1/2}$. The consumption of wine allocated to each worker -C_1 and C_2 respectively- is determined by the *Proportional Sharing*

[1]Another possibility -not considered in this paper- is that sabotage takes the form of offering information that, if taken into consideration, destroys part of the output (adverse selection).

[2]We do not deny that sabotage is, sometimes, an irrational action, taken by revenge, etc. Our purpose here is to show that even if such unfriendly feelings do not exist, there is room for sabotage.

Rule, i.e.

$$C_i = \frac{R_i}{R_1 + R_2}(R_1 + R_2)^{1/2},\ i = 1, 2.$$

For future reference we notice that this sharing rule is meritocratic, in the sense that allocates wine depending on relative inputs. Suppose that when the working day is about to finish, $R_1 = R_2 = 50$. Thus, $Y = 10$, $C_1 = C_2 = 5$. Now an unexpected event forces Beth to leave. Choices for Andy are to remain faithfully devoted to his own work, in which case he would obtain 21 extra units of grape or to destroy the crop assembled by Beth and pretend that somebody stole it.[3] In the first case his consumption of wine -the only thing he cares about- is

$$C_1 = \frac{71}{121}(121)^{1/2} \simeq 6,45.$$

In the second case, Andy's consumption of wine is

$$C_1 = \frac{50}{50}(50)^{1/2} \simeq 7,07.$$

Therefore, if Andy is rational, he will destroy Beth's crop. Suppose now that the sharing rule is *Egalitarian*, i.e.

$$C_i = \frac{(R_1 + R_2)^{1/2}}{2},\ i = 1, 2.$$

We notice that this rule is not meritocratic at all, in the sense that allocates wine irrespectively of relative inputs. In this case, faithful work yields to Andy $C_1 = 5.5$ and sabotage $C_1 = 3.5$, i.e. sabotage is not a rational action.[4]

What is going on in these examples? When an agent decides to sabotage other agent's crop, there are two effects. On the one hand wine output falls and there is

[3] The fact that Andy's output still 50 at the end of the day can be explained by saying that he spent the remaining labor time chasing the robber.

[4] We have not considered the possibility of stealing the output. In the example this is has been motivated by assuming that the colour of the grapes collected by the two agents are different.

less to distribute. This is bad from the saboteur's interest. On the other hand, the relative ranking of the agent whose crop has been sabotaged falls. This penalizes this agent and, with a given wine output, is good for the saboteur. Intuitively, the importance of the second effect depends on how meritocratic the sharing rule is, for instance in the egalitarian sharing rule this effect does not exist. When the rule is very meritocratic, the second effect dominates and sabotage may be rational, as in the case of the proportional sharing rule. The relationship between meritocracy and sabotage will be a central theme throughout our paper. Our analysis will unveil other determinants of sabotage too.

The model is presented in Section 2. In order to simplify the picture we make a number of simplifications. First we assume that the total quantity of labor supplied by any agent is fixed. Thus, labor can be spent on the production of an intermediate input (by exerting effort) or on the destruction of the inputs of other agents (sabotage). This assumption is made in order to focus attention on the choice between productive and sabotage activities. It is appropriated when length of working time is fixed exogenously by law, custom, etc. Second we assume that inputs are homogeneous. Thus total output depends on the sum of inputs. Third, the elasticity of output with respect to inputs is constant. This elasticity can be interpreted as an inverse measure of the degree of congestion.

In Section 3 we present a necessary condition to prevent sabotage. We define the degree of meritocracy of a sharing rule as the elasticity of the share of i in the output with respect to the inputs provided by i. For instance in the egalitarian rule this elasticity is zero and in the proportional rule this elasticity is $\frac{(n-1)}{n}$ where n is the number of agents. Let M be the marginal rate of substitution between sabotage and productive activities. M is a measure of how powerful destruction activities are in relation to production activities. The necessary condition for inexistence of sabotage has two parts: Either $M \leq n-1$ and then any sharing rule qualifies or $M > n-1$, and in this case the degree of meritocracy of a sharing rule must be less than a number which depends on M and the elasticity of output with respect to total inputs. When the number of agents and the technology of sabotage are variable and they might take any value, the latter condition simplifies and just says that the degree of meritocracy should be less than the elasticity of output with respect to the inputs.

In Section 4 we investigate the issue of the existence of a Nash equilibrium. We assume that sharing rules are anonymous, satisfy a separability assumption ($A1$) and a technical assumption involving the sign of certain derivatives ($A2$). Assumptions $A1-2$ are satisfied by most sharing rules used in the literature. We will keep them throughout the paper. We show that if M is smaller than one, a Nash equilibrium with zero sabotage exists, and if there are two agents -or more than two and the sharing rule satisfies a strong separability assumption ($A1'$)- zero sabotage is the unique Nash equilibrium (Proposition 1). However, if $A1'$ does not hold, there are cases in which a Nash Equilibrium with sabotage exists (Example 1). If M is larger than one there are always Nash equilibria in which all or all minus one agents do not supply a positive quantity of input (Remark 1). In fact, we present an example in which the necessary condition for the inexistence of sabotage holds but the only Nash equilibria are those described in Remark 1 (Example 2). The sharing rule used in this example, again, satisfies $A1$ but not $A1'$. Under the latter assumption we prove that if $M \leq n-1$ (so the first part of the necessary condition holds) there is a Nash equilibrium with zero sabotage (Proposition 2). Finally, we deal with the case where $M > n-1$. In this case the technology of destruction is very powerful and we need a new assumption ($A3$) that limits the degree of meritocracy of the sharing rule. Assumption $A3$ is just a generalization of the second part of the necessary condition for no sabotage to non symmetric allocations. For the class of sharing rules satisfying $A1' - 3$, a Nash equilibrium with zero sabotage exists (Proposition 3). Summing up for any parameters, there is a Nash equilibrium and for conditions that are slightly stronger than the necessary condition and for a somewhat restricted set of admissible sharing rules, there is a Nash equilibrium with zero sabotage. Notice that certain forms of sharing rules (those satisfying $A1$ but not $A1'$) are more prone to yield sabotage in a Nash equilibrium than others (those satisfying $A1'$) as it is shown in Examples 1 and 2.

In our model, since labor supply is fixed, zero sabotage implies efficiency. The reader may wonder why in our case efficiency may arise in equilibrium: Holmstrom (1982) has shown that if the whole output has to be distributed and individual efforts are not contractible, any Nash equilibrium of a game where efforts are strategies leads to an inefficient outcome. But in our model agents produce an intermediate input that is contractible. As it has been shown by Nandeibam (2002) in a model where sabotage is not possible, efficiency might arise in equilibrium in a team model if there are contractible intermediate inputs. Thus, our's and Nandeibam's results show the importance of contractible intermediate inputs.

Let us now comment on other papers dealing with sabotage. In the case of a capitalistic firm, Lazear (1989) was the first to point out that, in presence of sabotage, large differences in salaries become dysfunctional. In his model, agents are paid according to the position achieved in a contest. Holmstrom and Milgrom (1991) made the point that when a task is remunerated but other is not, incentives must be "low powered". Using a Principal-Agent model they showed that if compensation is linear in output, optimal compensation is flat on output.

This literature is surveyed in Gibbons (1998) and Prendergast (1999). In contrast with these models we assume that compensation can take a much more general form, that of a sharing rule. This creates difficulties to show the existence of a Nash equilibrium (in the case of a tournament or when compensation is linear, existence of a Nash equilibrium easily obtains).[5] In a model of Rent-Seeking, Konrad (2000) considered that the effort of an agent reduces rival´s performance by sabotaging their activities. His main result is that, in equilibrium, sabotage disappears with a large number of agents.

Summing up, our paper finds that the design of sharing rules is constrained by the presence of potential sabotage which, in turn, depends on five factors: (1) The degree of meritocracy (as in Lazear et alia), (2) the number of agents (as in Konrad), (3) the degree of congestion, (4) the technology of destruction and (5) the form of the sharing rule. Thus our analysis of the necessary and sufficient conditions for absence of sabotage has produced a picture that is considerably more complex than the one we had before.

2. The Model

The setting is one of cooperative production, see, e.g., Roemer and Silvestre (1993) for examples and applications. There are n agents. The input provided by agent i is denoted by $R_i \in \mathbb{R}_+$. A production function relates total output, Y, and the sum of inputs. This function is assumed to be of constant elasticity with decreasing returns to scale, i.e. $Y = (\sum_{i=1}^{n} R_i)^r$, where $0 < r < 1$. The parameter

[5] Itoh (1991) and Macho-Stadler and Pérez-Castrillo (1993) have analyzed the related case where cooperation among agents is desirable and give conditions for the existence of an equilibrium.

r is the elasticity of output with respect to total inputs and it can be interpreted as the inverse of the degree of congestion. The limit case $r = 1$, is uninteresting because in this case each agent is not constrained by the choices made by other agents. Total output is shared among agents by means of a sharing rule, i.e. a list of functions $S_i : \mathbb{R}^n_+ \to \mathbb{R}_+$ $i = 1, ..., n$ such that if C_i denotes the consumption of i,

$$C_i = S_i(R_1, .., R_n) \text{ and } \sum_{i=1}^{n} S_i(R_1, .., R_n) = Y \text{ for all } (R_1, .., R_n) \in \mathbb{R}^n_+$$

where S_i is a $\mathcal{C}^1$ function in $\mathbb{R}^n_{++}$. Sharing rules can also be written as

$$C_i = s_i(R_1, .., R_n)Y, \text{ with } s_i(R_1, .., R_n) \geq 0$$
$$\text{and} \sum_{i=1}^{n} s_i(R_1, .., R_n) = 1 \text{ for all } (R_1, .., R_n) \in \mathbb{R}^n_+$$

We will assume that $s_i(\ )$ is non decreasing on R_i, $s_i(R_1, .., R_n) > 0$ if $R_i > 0$, and that sharing rules are anonymous in the following sense:

$$\{R_1 = R_2 = ... = R_n\} \Rightarrow \{s_i = s_j \text{ and } \frac{\partial s_i}{\partial R_j} = \frac{\partial s_j}{\partial R_i}\ \forall i, j \text{ with } i \neq j\}.$$

This assumption holds if, for example, $s_i(R_1, .., R_n) = s(R_i, \sum_{k=1}^{n} f(R_k))$, i.e. if the share of i is a function independent of i, which depends on the input of i and the sum of a function of inputs. Examples of sharing rules fulfilling these conditions are (see Moulin (1987) and Pfingsten (1991)):

$$C_i = \frac{Y}{n},\ i = 1, ..., n \text{ (Egalitarian)}. \tag{2.1}$$

$$C_i = \frac{R_i}{\sum_{k=1}^{n} R_k}\ Y,\ i = 1, ..., n \text{ (Proportional)}. \tag{2.2}$$

$$C_i = \frac{(R_i)^\lambda}{\sum_{k=1}^{n} (R_k)^\lambda}\ Y,\ \lambda \in [0, 1],\ i = 1, ..., n. \tag{2.3}$$

$$C_i = g(\sum_{k=1}^{n} R_k)R_i + (1/n)(Y - g(\sum_{k=1}^{n} R_k)\sum_{k=1}^{n} R_k)), \ i = 1, ..., n, \tag{2.4}$$

$$\text{where } 0 \leq g(\sum_{k=1}^{n} R_k) \leq \frac{Y}{\sum_{k=1}^{n} R_k} \text{ for all } (R_1, .., R_n) \in \mathbb{R}_+^n.$$

In (2.3) we have described a class of sharing rules parametrized by λ. If $\lambda = 0$ we get the egalitarian sharing rule and if $\lambda = 1$, we have the proportional sharing rule. Thus, λ can be interpreted as a measure of meritocracy. In (2.4) the consumption of an agent is made of a payment to the input provided by the agent and an equal share on profits. If $g(\sum_{k=1}^{n} R_k)) = 0$ we get the egalitarian sharing rule, and if $g(\sum_{k=1}^{n} R_k) = \frac{Y}{\sum_{k=1}^{n} R_k}$ we have the proportional sharing rule. The function g is a measure of how relative effort is valued and thus measures the degree of meritocracy. A particular subclass of (2.4) is defined by $g(\sum_{k=1}^{n} R_k) = \frac{\alpha Y}{\sum_{k=1}^{n} R_k}$ with $0 \leq \alpha \leq 1$. For this particular function g we get a convex combination of the proportional and the egalitarian sharing rules.

$$C_i = (\frac{\alpha R_i}{\sum_{k=1}^{n} R_k} + \frac{1-\alpha}{n})Y, \ \alpha \in [0, 1], \ i = 1, ..., n. \tag{2.5}$$

The parameter α serves as a measure of the degree of meritocracy. Another subclass of (2.4) is given by $g(\sum_{k=1}^{n} R_k) = \mu r(\sum_{k=1}^{n} R_k)^{r-1}$ with $0 \leq \mu \leq \frac{1}{r}$.

$$C_i = \mu r(\sum_{k=1}^{n} R_k)^{r-1} R_i + \frac{1}{n}(\sum_{k=1}^{n} R_k)^r (1 - \mu r), \ \mu \in [0, \frac{1}{r}], \ i = 1, .., n. \tag{2.6}$$

For $\mu = 1$, each agent is paid according to her marginal productivity plus and an equal share of the surplus. Again, the parameter μ serves as a measure of the degree of meritocracy. For $\mu = 0$ we get the egalitarian sharing rule and for $\mu = \frac{1}{r}$ the proportional sharing rule.

Agents care only about their own consumption. As we remarked in the Introduction, the quantity of labor time is fixed. An agent, say i, can divide her working time, denoted by T, between productive labor, denoted by l_i^P and sabotage activities.[6] Let l_{ij} be the quantity of labor allocated by i to sabotage the input of agent j. The time constraint reads, $T = l_i^P + \sum_{j \neq i} l_{ij}$. The input provided by agent i depends on her own productive effort and the amount of time devoted by the remaining agents to sabotage the input of i, i.e.

$$R_i = R(l_i^P, l_{1i}, ...l_{(i-1)i}, l_{(i+1)i}, ...l_{ni})$$

where $R(\ )$ is a $\mathcal{C}^1$ function such that

$$\frac{\partial R}{\partial l_i^P} > 0 \text{ and } \frac{\partial R}{\partial l_{ji}} < 0. \tag{2.7}$$

We will assume that for each agent, say i, the strategic variable is the time devoted to sabotage activities, i.e. the vector $(l_{i1}, l_{i2}, l_{i(i-1)}, l_{i(i+1)}, l_{in})$. Time devoted to productive activities is determined by the constraint $l_i^P = T - \sum_{j \neq i} l_{ij}$. Thus, the decision problem for an agent is to choose the vector of sabotage activities in order to maximize $s_i(R_1, ..., R_n)(\sum_{k=1}^n R_k)^r$ subject to

$$R_j = R(T - \sum_{j \neq i} l_{ji}, l_{1j}, ...l_{(j-1)j}, l_{(j+1)j}, ...l_{nj}),\ j = 1, ..., n,$$

and taking the choice of sabotage activities of the rest of the agents as given. In other words, we look for a Nash Equilibrium (NE) in which the strategies are sabotage activities.

We postpone to Section 4 the problem of existence of NE. In Section 3 below we concentrate on the implications of guaranteeing that no agent has incentives to engage in sabotage.

[6] We are assuming that there is no possible protection against sabotage. This is only made by simplicity. It can be shown that identical results to those presented here hold if agents can devote part of their time to prevent sabotage by others.

3. A Necessary Condition for No Sabotage

If no agent has incentive to sabotage if all other agents do not sabotage, it must be that $\forall i, j$:

$$\frac{\partial C_i}{\partial l_{ij}} = (\sum_{k=1}^{n} R_k)^r (\frac{\partial s_i}{\partial R_j}\frac{\partial R_j}{\partial l_{ij}} - \frac{\partial s_i}{\partial R_i}\frac{\partial R_i}{\partial l_i^P}) + s_i r(\sum_{k=1}^{n} R_k)^{r-1}(\frac{\partial R_j}{\partial l_{ij}} - \frac{\partial R_i}{\partial l_i^P}) \leq 0, \tag{3.1}$$

where the partial derivatives are evaluated at the point where all working time is devoted to productive activities. This is a minimal requirement because if it is not fulfilled, all NE imply sabotage. We will see that such a minimal requirement imposes strong conditions on the form of the sharing rule and that such condition boilds down to a simple formula. Let

$$M = -\frac{\frac{\partial R_j}{\partial l_{ij}}}{\frac{\partial R_i}{\partial l_i^P}}$$

evaluated at the point where all working time is devoted to productive activities. M is the marginal rate of substitution between sabotage and productive activities.[7] From our assumptions it follows that $M > 0$. Now using the definition of M and dividing by $(\sum_{k=1}^{n} R_k)^r$, equation (3.1) above reads

$$-\frac{\partial s_i}{\partial R_j}M - \frac{\partial s_i}{\partial R_i} + \frac{s_i r}{\sum_{k=1}^{n} R_k}(-M-1) \leq 0.$$

Differentiating $\sum_{i=1}^{n} s_i(R_1, ..., R_n) = 1$ and using our anonymity assumption,

$$\frac{\partial s_i}{\partial R_i} = -(n-1)\frac{\partial s_j}{\partial R_i} = -(n-1)\frac{\partial s_i}{\partial R_j}. \text{ Thus,}$$

$$\frac{s_i r(M+1)}{nR_i} \geq \frac{\partial s_i}{\partial R_i}(\frac{M-n+1}{n-1}). \tag{3.2}$$

[7] Notice that because our symmetry assumption on $R(\ )$, $M = -\frac{\partial R_r}{\partial l_{ir}^S} / \frac{\partial R_i}{\partial l_i^P}$ any $r \neq i$.

Define

$$\epsilon = \frac{\partial s_i}{\partial R_i}\frac{R_i}{s_i}$$

as the elasticity of the share of i with respect to the input of i, evaluated at the point of zero sabotage (this elasticity will be interpreted below). From (3.1),

$$\frac{r(M+1)}{n} \geq \epsilon(\frac{M-n+1}{n-1}). \tag{3.3}$$

Now we have two cases. If $M \leq n-1$, the inequality (3.3) always holds. In words, if the marginal rate of substitution between sabotage and productive activities is small in relationship with the number of agents, no agent has incentives to start sabotage activities no matter how meritocratic the sharing rule is. However, if $M > n-1$ the above inequality implies that

$$\epsilon \leq \frac{r(M+1)(n-1)}{n(M-n+1)}. \tag{3.4}$$

In this case, ϵ is bounded. When the function $R(\ )$ can vary, M can take any value and the equation above has to hold for any value of M, the above equation reads

$$\epsilon \leq \frac{r(n-1)}{n}, \tag{3.5}$$

which in turn implies $\epsilon < r$.

The elasticity, ϵ is the degree of responsiveness of the sharing rule to the input provided by an agent. Thus it measures how meritocratic a sharing rule is. It can be easily checked that the measures of meritocracy discussed above in sharing rules (2.3)-(2.6) are related to ϵ, i.e.

$$\epsilon = \frac{\lambda(n-1)}{n},\ \epsilon = \frac{R_i g(nR_i)(n-1)}{(nR_i)^r},\quad \epsilon = \frac{\alpha(n-1)}{n},\ \epsilon = \frac{r\mu(n-1)}{n}$$

respectively. In these cases the necessary condition (3.5 above) reads as follows:

$$\lambda \leq r, \; g(nR_i) \leq \frac{r(nR_i)^r}{nR_i}, \; \alpha \leq r, \; \mu \leq 1.$$

In all the above cases the degree of meritocracy is bounded. In particular, the proportional sharing rule never satisfies this condition. Furthermore, for the class of sharing rules (2.6) the condition tells us that the payment to the input provided by an agent can not be above marginal productivity.

In Figure 1 we have pictured the region in the (ϵ, M) plane for which the necessary condition holds.

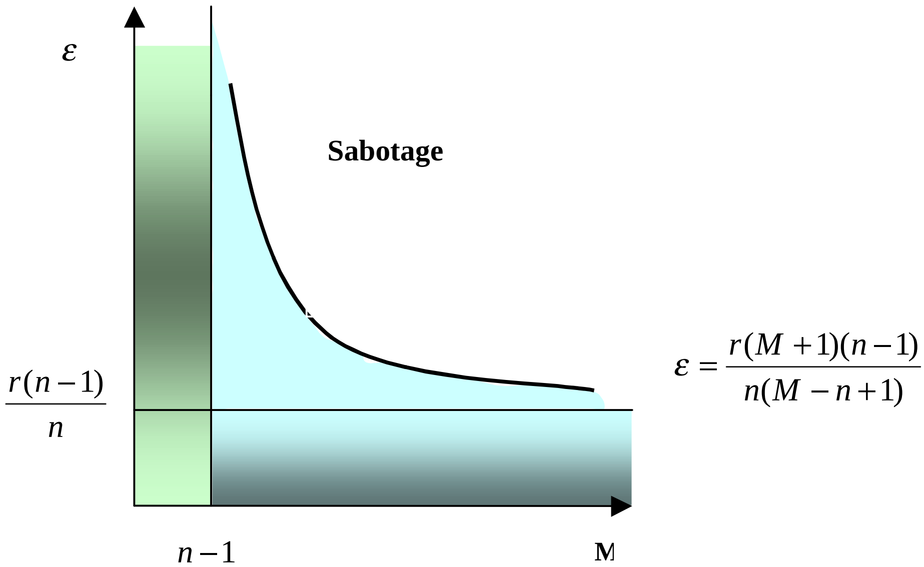


Figure 1

Notice the following:

1: An increase in r shifts the function $\frac{r(M+1)(n-1)}{n(M-n+1)}$ to the right and expands the area for which the necessary condition holds. This is explained by the fact that if congestion is high -r is small- aggregate output is not sensitive to inputs. Therefore, sabotage affects essentially to the shares s_i's, and a decrease in somebody' shares means an increase in everybody's else shares. Conversely with low congestion, returns from sabotage are small.

2: An increase in the number of agents shifts the function $\frac{r(M+1)(n-1)}{n(M-n+1)}$ to the right and expands the area for which the necessary condition holds. This is explained by the fact that with a large n, the share allocated to, say i, depends very little on the sum of $R_j's$. Thus the only effect of sabotage is through R_i.

Summing up, the necessary condition says that sabotage is likely in meritocratic organizations where the number of agents (n) is small, in which the technology of destruction (M) is very productive and in which the degree of congestion ($\frac{1}{r}$) is large.

4. Existence of Nash Equilibrium

In this section we study under what conditions a Nash equilibrium exists. We assume that the individual's input is given by

$$R_i = \max(T - \sum_{j \neq i} l_{ij} - M \sum_{j \neq i} l_{ji}, 0), \ i \in \{1, .., n\}.$$

Let us add the following assumptions:

$A1$. For all $i \in \{1, .., n\}$, $s_i = s(x_i, y)$ with $x_i = R_i$, $y = \sum_{k=1}^{n} f(R_k)$ and f a non decreasing and concave function.

$A2$. $\frac{\partial s}{\partial y} \leq 0$, $\frac{\partial^2 s}{\partial x_i \partial y} \leq 0$.

Assumption $A1$ says that the share allocated to i depends on the input supplied by i and an additively separable function of the inputs supplied by all agents. It is a special case of the anonymity assumption introduced in Section 2. Notice that sharing rules (2.3), (2.5) and (2.6) satisfy $A1-2$. Sometimes we will use a stronger version of $A1$, namely:

$A1'$. For all $i \in \{1,..,n\}$, $s_i = s(x_i, y)$ where $x_i = R_i$ and $y = \sum_{k=1}^{n} R_k$.

Assumption $A1'$ excludes sharing rules like (2.3).

Proposition 1. *Under assumptions $A1 - A2$ we get that:*
(i) if $M \leq 1$, zero sabotage for all i is a Nash equilibrium.
(ii) if $M < 1$ and $n = 2$, zero sabotage for all i is the unique Nash equilibrium.
(iii) if $M < 1$, $n > 2$, and $A1'$ holds, zero sabotage for all i is the unique Nash equilibrium.

Proof. (i) Let us see that if $l_{ji} = 0$, then the best response for agent i is $l_{ij} = 0$. For that it is enough to prove that $\frac{\partial C_i}{\partial l_{ij}} \leq 0$ for all l_{ij}. If we prove that $\frac{\partial s_i}{\partial l_{ij}} \leq 0$ for all l_{ij} we are done (by (3.1)). We compute

$$\frac{\partial s_i}{\partial l_{ij}} = (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_i}{\partial R_j}].$$

Since $s_i = s(x_i, y)$ where $x_i = R_i$ and $y = \sum_{k=1}^{n} f(R_k)$, we have that:

$$\frac{\partial s_i}{\partial R_j} = \frac{\partial s(R_i, y)}{\partial y}\frac{\partial y}{\partial R_j}.$$

Where $\frac{\partial y}{\partial R_j} = f'(R_j)$. Notice that $R_j = T - Ml_{ij}$, and $R_i = T - l_{ij}$. Since $M \leq 1$, $R_i \leq R_j$. Since f is concave, $f'(R_i) \geq f'(R_j)$, and since $\frac{\partial s}{\partial y} \leq 0$ and $\frac{\partial^2 s}{\partial x_i \partial y} \leq 0$ we get the following:

$$\frac{\partial s_i}{\partial R_j} = \frac{\partial s(R_i, y)}{\partial y}\frac{\partial y}{\partial R_j} = \frac{\partial s(R_i, y)}{\partial y} f'(R_j) \geq \frac{\partial s(R_j, y)}{\partial y} f'(R_i) = \frac{\partial s_j}{\partial R_i}.$$

Thus,

$$\frac{\partial s_i}{\partial l_{ij}} = (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_i}{\partial R_j}] \leq (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_j}{\partial R_i}].$$

Since $\sum_{k=1}^{n} s_k = 1$, then $\sum_{k=1}^{n} \frac{\partial s_k}{\partial R_i} = 0$, therefore,

$$\frac{\partial s_i}{\partial R_i} + \frac{\partial s_j}{\partial R_i} = -\sum_{k \neq i, k \neq j} \frac{\partial s_k}{\partial R_i} \geq 0.$$

which implies that

$$\frac{\partial s_i}{\partial l_{ij}} \leq (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_j}{\partial R_i}] \leq (-1)[\frac{\partial s_i}{\partial R_i} + \frac{\partial s_j}{\partial R_i}] \leq 0.$$

(ii) Let us show that if $M < 1$ and $n = 2$, this equilibrium is unique. Suppose we have a Nash equilibrium with positive sabotage (l^*_{ij}, l^*_{ji}). Since we have proved that the best response to zero sabotage is zero sabotage, in a Nash equilibrium with positive sabotage, $l^*_{ij} > 0$, and $l^*_{ji} > 0$. Since $M < 1$, we cannot have a Nash equilibrium with $R_1 = R_2 = 0$. Let R^*_i and R^*_j the individual's input at the equilibrium. Suppose that $R^*_i \leq R^*_j$. Then, $C_i(l^*_{ij}, l^*_{ji}) \leq \frac{1}{2}(R^*_i + R^*_j)^r$. If agent i reduces the sabotage activities to $l_{ij} = 0$, $R_i = T - Ml^*_{ji}$ and $R_j = T - l^*_{ji}$. Then, $R_i > R_j$ and the total output increases. Thus, $C_i(0, l^*_{ji}) \geq \frac{1}{2}(R_i + R_j)^r > \frac{1}{2}(R^*_i + R^*_j)^r \geq C_i(l^*_{ij}, l^*_{ji})$. So, we can not have an equilibrium with positive sabotage.

(iii) Let us see that if $s_i = s(R_i, \sum_j R_j)$ we can also guarantee that the equilibrium is unique for any number of agents. Suppose we have an equilibrium with positive sabotage. Then,

Step 1. There is at least one agent i such that $\sum_{j=1,j\neq i}^{n} l_{ji} \leq T$.

Suppose that for all agent i $\sum_{j=1,j\neq i}^{n} l_{ji} > T$. then, if we sum for all agents, $\sum_i \sum_{j=1,j\neq i}^{n} l_{ji} > nT$, but this is impossible since, by the time constraint, for all j, $\sum_{i=1,i\neq j}^{n} l_{ji} \leq T$.

Step 2. There is at least one agent i such that $R_i > 0$.

Suppose that, for all agent j , $R_j = 0$. Then $C_j = 0$ for all j. By Step 1 we know that there is an agent i such that $\sum_{j=1,j\neq i}^{n} l_{ji} \leq T$. Since $M < 1$, and $R_i = 0$, the amount of time devoted to sabotage activities by this agent i should be strictly positive. But this can not be an equilibrium. If this agent reduce her sabotage activities, the total output will be positive and her input positive. Consequently, she will get a positive amount. Therefore, she will be better off.

Step 3. There is no an equilibrium with positive sabotage such that $R_1 = R_2 =$

... $= R_n$.

Suppose on the contrary that such an equilibrium exist. Then by the anonymity condition, $C_i = \frac{1}{n}(\sum_j R_j)^r$. Suppose agent i reduces her sabotage activity to zero. For each j let $\hat{R}_j$ be the input in the new situation. Since $M < 1$, $\hat{R}_i > \hat{R}_j$ for all $j \neq i$. Thus, $\hat{C}_i \geq \frac{1}{n}(\sum_j \hat{R}_j)^r > \frac{1}{n}(\sum_j R_j)^r$, which implies that agent i is better off.

Step 4. In an equilibrium with positive sabotage, if $R_i \leq R_j$ then $l_{ij} = 0$.

Suppose on the contrary that $l_{ij} > 0$. Let us see that this can not be an equilibrium because $\frac{\partial s_i}{\partial l_{ij}} < 0$.

$$\frac{\partial s_i}{\partial l_{ij}} = (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_i}{\partial R_j}].$$

Since $s_i = s(R_i, y)$ with $y = \sum_j R_j$, we have that:

$$\frac{\partial s_i}{\partial R_j} = \frac{\partial s(R_i, y)}{\partial y}.$$

Since $R_i \leq R_j$ and $\frac{\partial^2 s}{\partial x_i \partial y} \leq 0$ we get the following:

$$\frac{\partial s(R_i, y)}{\partial y} \geq \frac{\partial s(R_j, y)}{\partial y} = \frac{\partial s_j}{\partial R_i}.$$

Thus,

$$\frac{\partial s_i}{\partial l_{ij}} = (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_i}{\partial R_j}] \leq (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_j}{\partial R_i}].$$

Since $\sum_{k=1}^n s_k = 1$, then $\sum_{k=1}^n \frac{\partial s_k}{\partial R_i} = 0$, therefore,

$$\frac{\partial s_i}{\partial R_i} + \frac{\partial s_j}{\partial R_i} = -\sum_{k \neq i, k \neq j} \frac{\partial s_k}{\partial R_i} \geq 0.$$

which, since $M < 1$, implies that

$$\frac{\partial s_i}{\partial l_{ij}} \leq (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_j}{\partial R_i}] < (-1)[\frac{\partial s_i}{\partial R_i} + \frac{\partial s_j}{\partial R_i}] \leq 0.$$

Step 5. Suppose that, in an equilibrium with positive sabotage, $R_1 \leq R_2 \leq ... \leq R_n$ with a strict inequality because Step 3. Then if $R_j \leq R_i$ and $i \neq n$, $l_{ij} = 0$. Suppose on the contrary that $l_{ij} > 0$. Let us see that this can not be an equilibrium because $\frac{\partial s_i}{\partial l_{ij}} < 0$.

$$\frac{\partial s_i}{\partial l_{ij}} = (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_i}{\partial R_j}].$$

Since $i \neq n$, $R_i \leq R_n$ and since $\frac{\partial^2 s}{\partial x_i \partial y} \leq 0$ we get that:

$$\frac{\partial s_i}{\partial R_j} = \frac{\partial s(R_i, y)}{\partial y} \geq \frac{\partial s(R_n, y)}{\partial y} = \frac{\partial s_n}{\partial R_i}.$$

As it was show in Step 4,

$$\frac{\partial s_i}{\partial R_i} + \frac{\partial s_n}{\partial R_i} = -\sum_{k \neq i, k \neq j} \frac{\partial s_k}{\partial R_i} \geq 0.$$

which, since $M < 1$, implies that

$$\frac{\partial s_i}{\partial l_{ij}} \leq (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_n}{\partial R_i}] < (-1)[\frac{\partial s_i}{\partial R_i} + \frac{\partial s_n}{\partial R_i}] \leq 0.$$

Step 6. There is no an equilibrium with positive sabotage.
By Step 4 and Step 5, we know that for all $i \neq n$, and for all j, $l_{ij} = 0$. Thus, if there is an equilibrium with positive sabotage, only agent n is using part of her time in sabotage activities. But since $M < 1$, and $R_n \geq R_i$ for all i, the productive activities of agent n is larger than the productive activities of any of the other agents. Which implies that we can not get an equilibrium with positive sabotage. ■

In the above proposition we have shown that in order to guarantee uniqueness we have to restrict the class of admissible sharing rules for $n > 2$. The following example shows that without such restriction, uniqueness no longer holds.

Example 1. *Let* $n = 4$ *and*

$$C_i = \frac{R_i^\lambda}{\sum R_j^\lambda}\left(\sum R_j\right)^r$$

Notice that this sharing rule is of the form $s_i(R_i, y)$ *where* $y = \sum_{j=1}^4 f(R_j)$ *with* f *non decreasing and strictly concave,* $\frac{\partial s_i}{\partial y} \leq 0$ *and* $\frac{\partial^2 s_i}{\partial y \partial x_i} \leq 0$*. Therefore* $A1-2$ *hold. Let* $\lambda = \frac{1}{4}$, $r = \frac{1}{10}$, $M = \frac{1}{2}$ *and* $T = 1$*. By Proposition 1, zero sabotage is a Nash equilibrium. Consider now the following strategy where agent 1 is sabotaged by all other agents, and she does not sabotage to any one. That is,* $l_{12} = l_{13} = l_{14} = 0$, $l_{23} = l_{24} = 0$, $l_{32} = l_{34} = 0$, $l_{42} = l_{43} = 0$, *and* $l_{21} = l_{31} = l_{41} = \frac{2}{3}$*. Given this strategy,* $R_1 = 0$, $R_2 = R_3 = R_4 = \frac{1}{3}$*. Thus* $C_1 = 0$, $C_2 = C_3 = C_4 = \frac{1}{3}$*. The proof that this strategy is a Nash equilibrium is left to the Appendix.*

Let us now consider the case where $M \geq 1$. In this case, there are two kinds of trivial Nash equilibria: In the first kind, no positive output is produced. In the second kind (only possible if $n > 2$), only one agent produces a positive input.

Remark 1. *Let* $M \geq 1$.
I) There is a Nash equilibrium such that $R_i = 0$ *for all* i *(Proof: For each* i*, let* $l_{ii+1} = T$ *(modulo* n*) and* $l_{ij} = 0$ *otherwise. Clearly* $R_i = 0$ *for all* i*. It is also clear that no agent can deviate profitably).*
II) If $n > 2$*, there is a Nash equilibrium such that there is a single agent* j *with* $R_j > 0$ *and* $R_i = 0$ *for all* $i \neq j$ *(Proof: wlog let* $j = n$*. For each* $i \neq n$*, let* $l_{ii+1} = T$ *(modulo* $n-1$*) and* $l_{ij} = 0$ *otherwise,* $l_{ni} = 0$ *for all* i*. Clearly* $R_i = 0$ *for all* $i \neq n$*. It is also clear that no agent can deviate profitably).*

The Nash equilibria in the remark above are not strict, because all but, at most one agent, are indifferent among *any* distribution of their time. Therefore these kind of equilibria are not very robust. However, in some cases they are the only Nash equilibria: The next example shows that if $n > 2$, for certain sharing rules fulfilling $A1$, but not $A1'$, even if $M \leq n-1$ (so the necessary condition for zero to be an equilibrium is satisfied), zero sabotage is not an equilibrium, and there is no an equilibrium where at least two agents produce a positive input.

Example 2. *Let $n = 3$ and*

$$C_i = \frac{R_i^\lambda}{\sum R_j^\lambda}\left(\sum R_j\right)^r$$

Let $\lambda = \frac{1}{2}$ and $r = 0.3$, $M = 2$ and $T = 1$. Then zero is not an equilibrium. Let us see for example that $l_{12} = l_{13} = 0$ is not a best response for agent 1 to $(l_{21}, l_{23}, l_{31}, l_{32}) = (0, 0, 0, 0)$. If agent one uses half of her time to sabotage agent two, she is better off than just working all the time. That is, $C_1(0, ..., 0) < C_1(0.5, 0, 0, 0, 0, 0)$.

$$C_1(0, ..., 0) = \frac{1}{3}(3)^{0.3} = .46346$$

$$C_1(0.5, 0, .., 0) = \frac{(0.5)^{0.5}}{(0.5)^{0.5} + 1}(1.5)^{0.3} = .46779$$

The proof that there is no equilibrium other than those described in Remarks I and II is left to the Appendix.

From the previous example, it is clear that equilibrium with no sabotage will not exists for certain sharing rules. Thus, in what follows we restrict to those sharing rules fulfilling $A1'$.

The following Lemma tell us that, under $A1'$ we can restrict our attention to symmetric best responses.

Lemma 1. *If* $l_i = (l_{i1}, ..., l_{ii-1}, l_{ii+1}, .., l_{in})$ *is a best response for agent* i *to* $l_{-i} = 0$*, then* $\hat{l}_i = (l, ..., l)$ *with* $l = \frac{\sum_{j=1, j\neq i}^{n} l_{ij}}{n-1}$ *is also a best response for agent* i *to* $l_{-i} = 0$.

Proof. Notice first that if l_i is a best response to $l_{-i} = 0$, then $T - Ml_{ij} \geq 0$ for all j. Suppose not, that is, suppose that there is an agent j such that $T - Ml_{ij} < 0$. Then agent i can decrease the time dedicated to sabotage agent j up to a point such that $T - Ml'_{ij} = 0$. Thus agent i will increase her output without affecting the output of the other agents, which implies that she will be better off and will contradict that l_i is a best response against $l_{-i} = 0$.
Since $T - Ml_{ij} \geq 0$ for all j, the total output of all agents but i is: $\sum_{j\neq i} R_j(l_i, l_{-i}) = (n-1)T - M \sum_{j=1, j\neq i}^{n} l_{ij}$. Notice also that $T - Ml \geq 0$, otherwise for the agent k such that $l_{ik} = \max_j l_{ij}$, $T - Ml_{ik} < 0$. Then, the output of all agents but i under $\hat{l}_i$ is $R_j(\hat{l}_i, l_i) = T - Ml$, and $\sum_{j\neq i} R_j(\hat{l}_i, l_{-i}) = (n-1)T - M(n-1)l = \sum_{j\neq i} R_j(l_i, l_{-i})$. Furthermore, $R_i(l_i, l_{-i}) = R_i(\hat{l}_i, l_{-i})$. Therefore $s_i(l_i, l_{-i}) = s_i(\hat{l}_i, l_{-i})$ which implies that $C_i(l_i, l_{-i}) = C_i(\hat{l}_i, l_{-i})$. ■

Next Proposition shows that, under $A1' - 2$, if the marginal rate of substitution between sabotage and productive activities, M, is less than $n-1$, (so the first part of the necessary condition obtained in the previous section holds) zero sabotage is a Nash equilibrium. The intuition is that in this case the damage that agents can inflict each other is small.

Proposition 2. *Under assumptions* $A1', A2$ *and if* $M \leq n-1$, $l_{ij} = 0$ *for all* $i, j \in \{1, .., n\}$ *is a Nash equilibrium.*

Proof. Let us see that if $l_{-i} = 0$, the best response for agent i is $l_i = 0$. Because of the previous Lemma, if we prove that for all $j \neq i$, $\frac{\partial C_i(l_i, 0)}{\partial l_{ij}} \leq 0$ for all

$l_i = (l, ..., l)$, then $l_i = 0$ is the best response for this agent. We know that

$$\frac{\partial C_i}{\partial l_{ij}} = \left[\frac{\partial s_i}{\partial l_{ij}}\right] y^r + s_i r \left[-1 - M\right] y^{r-1}.$$

Therefore, $\frac{\partial C_i}{\partial l_{ij}} \leq 0$, whenever $\frac{\partial s_i}{\partial l_{ij}} \leq 0$, where,

$$\begin{aligned}\frac{\partial s_i}{\partial l_{ij}} &= \frac{\partial s_i}{\partial R_i}(\frac{\partial R_i}{\partial l_{ij}}) + \frac{\partial s_i}{\partial R_j}(\frac{\partial R_j}{\partial l_{ij}}) = \\ &\quad (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_i}{\partial R_j}].\end{aligned}$$

Since $s_i = s(x_i, y)$ where $x_i = R_i$ and $y = \sum_{k=1}^n R_k$, we have that:

$$\frac{\partial s_i}{\partial R_j} = \frac{\partial s_i}{\partial y}\frac{\partial y}{\partial R_j} = \frac{\partial s_i}{\partial y}.$$

Notice that for all vectors $l_i = (l, .., l)$ and because $M \leq n-1$, $R_i = T-(n-1)l \leq R_j = T - Ml$. By Assumption A2, $\frac{\partial^2 s}{\partial x_i \partial y} \leq 0$, we get $\frac{\partial s_i}{\partial R_j} = \frac{\partial s(R_i,y)}{\partial y} \geq \frac{\partial s(R_j,y)}{\partial y} = \frac{\partial s_j}{\partial R_i}$. Thus,

$$\frac{\partial s_i}{\partial l_{ij}} = (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_i}{\partial R_j}] \leq (-1)[\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_j}{\partial R_i}].$$

Since $\sum_{k=1}^n s_k = 1$, then $\sum_{k=1}^n \frac{\partial s_k}{\partial R_i} = 0$. Furthermore, we know that $\frac{\partial s_k}{\partial R_i} = \frac{\partial s_j}{\partial R_i}$ for all $j, k \neq i$. Therefore,

$$\frac{\partial s_i}{\partial R_i} + (n-1)\frac{\partial s_j}{\partial R_i} = 0.$$

Since $M \leq n-1$,

$$\frac{\partial s_i}{\partial R_i} + M\frac{\partial s_j}{\partial R_i} \geq \frac{\partial s_i}{\partial R_i} + (n-1)\frac{\partial s_j}{\partial R_i} = 0,$$

which implies that $\frac{\partial s_i}{\partial l_{ij}} \leq 0$ as we wanted to prove. ■

Finally, we study the case where $M > n-1$. In this case, the second part of the necessary condition obtained in the previous section suggests that we have to impose an additional condition in order to guarantee zero sabotage in equilibrium.

This condition is the following.

$A3$. For each agent i, and for $l_i = (l, .., l)$, $l \geq 0$, $l_{-i} = 0$, $-\frac{\partial s}{\partial y}\frac{y}{s_i} \leq r$.

Assumption $A3$ requires the elasticity of the share of i with respect to the total output evaluated at points where the sabotage activities of all agents but i is cero to be less than r. This condition, evaluated at the point of zero sabotage, coincides with the second part of the necessary condition, equation (3.5). To see this, notice that, in those points,

$$\frac{\partial s_i}{\partial R_i} = -(n-1)\frac{\partial s}{\partial y},$$

and since $y = nR_i$,

$$-\frac{\partial s}{\partial y}\frac{y}{s_i} = \frac{\partial s_i}{\partial R_i}\frac{nR_i}{(n-1)s_i} \leq r,$$

which implies that

$$\frac{\partial s_i}{\partial R_i}\frac{R_i}{s_i} \leq \frac{r(n-1)}{n}.$$

For sharing rules (2.5) and (2.6) Assumption $A3$ is satisfied if:

$$\alpha \leq \frac{r}{n},\ \mu \leq \frac{1}{n},$$

which imposes an extra restriction on the degree of meritocracy.

Proposition 3. *Under assumptions $A1', A2$ and $A3$ if $M > n-1$, $l_{ij} = 0$ for all $i, j \in \{1, .., n\}$ is a Nash equilibrium.*

Proof. Let us see that if $l_{-i} = 0$, the best response for agent i is $l_i = 0$. Because of the previous Lemma, if we prove that for all $j \neq i$, $\frac{\partial C_i(l_i,0)}{\partial l_{ij}} \leq 0$ for all $l_i = (l, ..., l)$, then $l_i = 0$ is the best response for this agent. We know that

$$\frac{\partial C_i}{\partial l_{ij}} = y^{r-1}[\frac{\partial s_i}{\partial l_{ij}}y + s_i r(-1-M)].$$

Under our assumptions, the derivative of the share of agent i with respect to the sabotage activity can be written as follows:

$$\frac{\partial s_i}{\partial l_{ij}} = (-1)\frac{\partial s_i}{\partial x_i} + \frac{\partial s_i}{\partial y}(-1 - M).$$

Furthermore, since $\frac{\partial s_i}{\partial R_i} \geq 0$, and $\frac{\partial s_i}{\partial R_i} = \frac{\partial s_i}{\partial x_i} + \frac{\partial s_i}{\partial y}$, we have that $(-1)\frac{\partial s_i}{\partial x_i} \leq \frac{\partial s_i}{\partial y}$. Thus,

$$\frac{\partial s_i}{\partial l_{ij}} \leq \frac{\partial s_i}{\partial y}(-M).$$

Since $-\frac{\partial s_i}{\partial y}\frac{y}{s_i} \leq r$,

$$\frac{\partial s_i}{\partial l_{ij}}y - s_i r(1+M) \leq -M\frac{\partial s_i}{\partial y}y - s_i r(1+M) \leq M s_i r - s_i r(1+M) = -s_i r \leq 0,$$

which implies that $\frac{\partial C_i(l_i,0)}{\partial l_{ij}} \leq 0$ as we wanted to prove. ■

5. Conclusion

The classical approach to cooperative production (see, e.g., Sen (1966), Holmstrom (1982) and Fabella (1988)) assumes that shirking may be a problem but agents never engage in sabotage activities. In this paper we have worked out the consequences of assuming that agents can sabotage other agent's inputs and we have studied necessary and sufficient conditions to avoid sabotage in a Nash equilibrium. Let us comment on some possible extensions of our work.

1. **Technology:** We have assumed that inputs are substitutes. The case in which inputs are complements deserves attention. In this case, one would expect less incentive to sabotage than in the case assumed in this paper. This would only reinforce one of the points of this paper, namely, that the technology is one of the main forces shaping the relationship between meritocracy and sabotage.

2. **Repeated Interaction:** One could argue that in the setting of cooperative production, agents interact repeatedly and thus, the right approach to model the strategic aspects is by means of a Supergame. Unfortunately, the theory of supergames predicts a large number of possible outcomes, specially if agents are sufficiently patient. However, if we restrict our attention to trigger strategies, (short-run) Nash equilibrium of the sort analyzed in this paper is very important because: i) it serves as the only credible threat and ii) payoffs achieved in any long-run Perfect Nash Equilibrium must be greater or equal than the payoffs achieved in the short-run Nash equilibrium (see Friedman (1971)). Uniqueness of short-run Nash equilibrium (or lack of it) is also important in finitely repeated games (see the discussion in Osborne and Rubinstein (1994), pp. 159-60).

3. **Anonymity and Symmetry.** Our results are restricted to anonymous or symmetric sharing rules, which are those customarily used in the literature (see e.g. Moulin (1987), Pfingsten, A. (1991) or Roemer and Silvestre (1993)). Our methods can be used to show that, in some cases, symmetry may be necessary if we want to avoid sabotage. For instance, take the generalized equal benefit sharing rule:

$$C_i = r(\sum_{j=1}^{n} R_j)^{r-1} R_i + \theta_i((\sum_{j=1}^{n} R_j)^r - r(\sum_{j=1}^{n} R_j)^{r-1} \sum_{j=1}^{n} R_j)$$

with $\sum_{j=1}^{n} \theta_i = 1$ and $\theta_i > 0$, $i = 1, 2, ..., n$. The first part of the formula above is the marginal productivity of an agent. The second part is the share of i (θ_i) in the profits of the firm. The condition of no sabotage reads:

$$\theta_i \geq \frac{R_i}{\sum_{j=1}^{n} R_j} - \frac{1}{(1-r)(1+M)}.$$

which for large M reads $\theta_i \geq \frac{R_i}{\sum_{j=1}^{n} R_j}$, for all i. Given that $\sum_{j=1}^{n} \theta_i = 1$, it must be that $\theta_i = \frac{R_i}{\sum_{j=1}^{n} R_j}$ i.e. the share in the profits of the firm must be equal to the output of i divided by the sum of inputs. If the quantity of inputs provided by each agent is the same, $\theta_i = 1/n$ for all i.

4. **Agents with Different Productivities:** An interesting extension of our work is to consider that agents have different productivities in both productive and sabotage activities. For instance assuming that agents can be of two types in each activity we have four types of agents. This model opens up new possibilities like the study of how sabotage depends on the relative number of types and on the relative productivity in production and sabotage.

5. **Characterization of the form taken by sabotage**: We have shown that in certain cases (first part of Remark 1) sabotage is symmetric in equilibrium. However, other results show that sabotage may be very asymmetric, i.e. it might be concentrated on some individuals (e.g. Example 1 or the second part of Remark 1). Also sabotage could be partial, with some resources devoted to sabotage and others to productive activities, (as in Example 1) or extreme with all resources devoted to sabotage (as in the first part of Remark 1). It would be nice to know more about what kind of sabotage we might expect in equilibrium and how the form of sabotage is shaped by factors like the form of the sharing rules, the technology of destruction, the number of agents, etc.

Summing up, in this paper we have studied a model of cooperative production and we have shown that an organization populated by rational agents might be self-destructive. In broad terms we might summarize our findings by saying that the conflict in an organization depends positively on the degree of congestion ($\frac{1}{r}$), the possibilities of destruction (M) and the degree of meritocracy (ϵ), and negatively on the number of agents (n). And that certain forms of sharing rules (e.g. (2.3)) are more prone than others to yield sabotage. We hope that our paper will stimulate more work in this area that will produce additional insights on this problem.

References

[1] Fabella, R.V. (1988). “Natural Team Sharing and Team Productivity,” *Economics Letters* 27, 105-110.

[2] Friedman, J. W. (1971). ”A Non-Cooperative Equilibrium for Supergames”, *Review of Economic Studies,* 38, 1-12

[3] Gibbons, R. (1998). “Incentives in Organizations,” *Journal of Economic Perspectives,* 12 (4), 115-132.

[4] Holmstrom, B. (1982). “Moral hazard and Teams,” *Bell Journal of Economics*, 13, 324-340.

[5] Holmstrom, B., and Milgrom, P. (1991). “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics and Organization,* 7, Special Issue, 24-52.

[6] Itoh, H. (1991). “Incentives to Help in Multi-Agent Situations,” *Econometrica,* 59 (3), 611-636.

[7] Konrad, K. A. (2000). “Sabotage in Rent-Seeking Contests,” *Journal of Law, Economics and Organizations*, 16 (1), 155-165.

[8] Lazear, E. P. (1989). “Pay Equality and Industrial Politics,” *Journal of Political Economy*, 97 (3), 561-580.

[9] Macho-Staedler, I., and Pérez-Castrillo, D. (1993). “Moral Hazard with Several Agents,” *International Journal of Industrial Organization*, 11, 73-100.

[10] Moulin, H. (1987). "Equal or proportional Division of Surplus and Other Methods," *International Journal of Game Theory*, 16, 161-186.

[11] Nandeibam, S. (2002). "Implementation in teams," Working paper, University of Birmingham.

[12] Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory.* The MIT Press.

[13] Prendergast, C. (1999). "The Provision of Incentives in Firms," *Journal of Economic Literature,* 37,7-63.

[14] Pfingsten, A. (1991). "Surplus Sharing Methods," *Mathematical Social Sciences*, 21, 287-301.

[15] Roemer, J., and Silvestre, J. (1993). "The Proportional Solution for Economies with Private and Public Ownership," *Journal of Economic Theory*, 59, 426-444.

6. Appendix.

The strategy described in Example 1 is a Nash equilibrium. Let us see that $l_1 = (0,0,0)$, $l_2 = (\frac{2}{3},0,0)$, $l_3 = (\frac{2}{3},0,0)$ and $l_4 = (\frac{2}{3},0,0)$ is a Nash equilibrium.

For this strategy $R_1 = 0$ and $C_1 = 0$. Therefore, agent 1 can not improve her payoff by increasing her sabotage activities. Thus, $l_1 = (0,0,0)$ is a best response for this agent against l_{-1}. Let us see that for each agent $i = 2,3,4$, $l_i = (\frac{2}{3},0,0)$ is a best response against l_{-i}. Let us prove this for agent 4. The same argument applies for agents 2 and 3.

Step 1. Let (l_{41}, l_{42}, l_{43}) be a best response against l_{-4}. Suppose, without loss of generality, that $R_1 \leq R_2 \leq R_3$. Then if $l_{42} > 0$, $R_1 = 0$.

Suppose that $R_1 > 0$. This can not be a best response because agent 4 can be better off using the strategy $(l_{41}+\varepsilon,\, l_{42}-\varepsilon,\, l_{43})$ with $\varepsilon > 0$. Notice that under this strategy the input of agent 1 will decrease in $M\varepsilon$, the input of agent 2 will increase $M\varepsilon$, the input of agent 4 will remain the same, and also the total output. Since f is strictly concave, $f(R_2+M\varepsilon) - f(R_2) < f(R_1) - f(R_1 - M\varepsilon)$. Let $y = f(R_1)+f(R_2)+f(R_3)+f(R_4)$, and $y' = f(R_2+M\varepsilon)+f(R_1-M\varepsilon)+f(R_3)+f(R_4)$. Since $\frac{\partial s_i}{\partial y} < 0$, $s_4(R_4, y) < s_4(R_4, y')$. Thus agent 4 is better off with this new strategy.

Step 2. Let (l_{41}, l_{42}, l_{43}) be a best response against l_{-4}. Suppose, without loss of generality, that $R_1 \leq R_2 \leq R_3$. Then $l_{42} = l_{43} = 0$.

Suppose that $l_{42} > 0$, then by Step 1, $R_1 = 0$. Which implies that $l_{41} = \frac{2}{3}$ (it can not be greater than $\frac{2}{3}$, otherwise this agent could decrease l_{41} without affecting the input of agent 1 and increasing the total output, being better off in this case). Furthermore, for this strategy, $R_4 < R_3$ and $R_4 < R_2$. Thus $C_4 < \frac{1}{3}(R_2+R_3+R_4)^r$. Suppose that agent 4 uses $(\frac{2}{3},0,0)$. Let R_1', R_2', R_3',

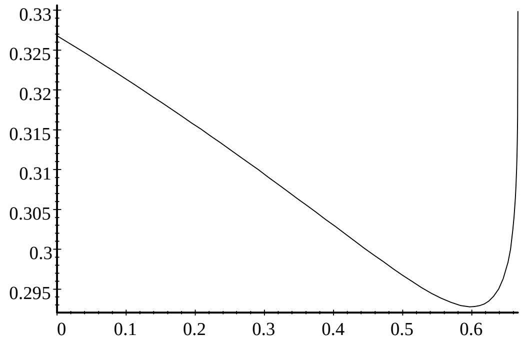
R'_4 the inputs for this strategy. Notice that $R'_2 = R'_3 = R'_4$, $R'_1 = 0$ and the total output has increased. Thus, $C'_4 = \frac{1}{3}(R'_2 + R'_3 + R'_4)^r > C_4$ which contradicts that (l_{41}, l_{42}, l_{43}) is a best response against l_{-4}.

Step 3. The strategy $(\frac{2}{3}, 0, 0)$ is a best response for agent 4 against l_{-4}.

By Step 2, if l_4 is a best response against l_{-4} and $R_1 \leq R_2 \leq R_3$, then $l_4 = (l_{41}, 0, 0)$ with $l_{41} \leq \frac{2}{3}$. For any strategy like l_4 we have: $R_1 = \frac{1}{3} - \frac{1}{2}l_{41}$, $R_2 = \frac{1}{3}$, $R_3 = \frac{1}{3}$, and $R_4 = 1 - l_{41}$. Thus,

$$C_4(l_{41}) = \frac{(1-l_{41})^{\frac{1}{4}}}{(1-l_{41})^{\frac{1}{4}} + 2(\frac{1}{3})^{\frac{1}{4}} + (\frac{1}{3} - \frac{1}{2}l_{41})^{\frac{1}{4}}}(2 - \frac{3}{2}l_{41})^{\frac{1}{10}}.$$

The graph of this function is represented in the following figure:



$C_4(l_{41})$

Where $C_4(0) = .326\,81$, and $C_4(\frac{2}{3}) = .333\,33$. Therefore, $C_4(l_{41}) < C_4(\frac{2}{3})$ for any $l_{41} < \frac{2}{3}$. ■

There is no equilibrium with two agents producing a positive input in Example 2. Suppose that there is an equilibrium with two agents producing a positive input. Since zero sabotage is not an equilibrium, any possible equilibria involve positive sabotage.

Step 1. There is no equilibrium with positive sabotage such that $R_1 = R_2 =$

$R_3 > 0$.

Suppose on the contrary that there is an equilibrium $((l_{12}, l_{13}), (l_{21}, l_{23}), (l_{31}, l_{32}))$ such that $R_1 = R_2 = R_3$. But, in those points, it is easy to see that because anonymity

$$\frac{\partial s_i}{\partial l_{ij}} = -\frac{\partial s_i}{\partial R_i} - 2\frac{\partial s_i}{\partial R_j} = -\frac{\partial s_i}{\partial R_i} - \frac{\partial s_j}{\partial R_i} - \frac{\partial s_k}{\partial R_i} = 0.$$

which implies $\frac{\partial C_i}{\partial l_{ij}} < 0$, i.e. agent i is better off reducing her sabotage activities.

Step 2. In equilibrium, if $R_i = 0$ for some i, then $l_{ij} = l_{ik} = 0$.

First notice that in equilibrium if $R_i = 0$, then $T - l_{ij} - l_{ik} - 2(l_{ji} + l_{ki}) = 0$. This is because, by assumption, R_j and R_k are positive and if $T - l_{ij} - l_{ik} - 2(l_{ji} + l_{ki}) < 0$ agent j or k can reduce her sabotage activity against agent i affecting only her input which makes her better off. Then, $l_{ij} = l_{ik} = 0$, because otherwise agent i decreasing sabotage, increases her input and gets a positive share.

Step 3. Suppose we have an equilibrium such that $R_j \leq R_k$. Let (l_{ij}, l_{ik}) the strategy of agent i. If $l_{ik} > 0$, then $R_j = 0$.

Suppose that $R_j > 0$. This can not be an equilibrium because agent i can be better off using the strategy $(l_{ij} + \varepsilon,\ l_{ik} - \varepsilon)$ with $\varepsilon > 0$. Notice that under this strategy the input of agent j will decrease in $M\varepsilon$, the input of agent k will increase $M\varepsilon$, the input of agent i will remain the same, and also the total output. Since f is strictly concave, $f(R_k + M\varepsilon) - f(R_k) < f(R_j) - f(R_j - M\varepsilon)$. Let $y = f(R_j) + f(R_k) + f(R_i)$, and $y' = f(R_k + M\varepsilon) + f(R_j - M\varepsilon) + f(R_i)$. Since $\frac{\partial s_i}{\partial y} < 0$, $s_i(R_i, y) < s_i(R_i, y')$. Thus agent i is better off with this new strategy.

Step 4. Suppose that $((l_{12}, l_{13}), (l_{21}, l_{23}), (l_{31}, l_{32}))$ is an equilibrium such that $R_1 \leq R_2 \leq R_3$ with at least a strict inequality (this follows from Step 1). Then $l_{12} = l_{13} = 0$.

First we will see that, at this point, $\frac{\partial s_1}{\partial l_{12}} < 0$,

$$\frac{\partial s_1}{\partial l_{12}} = (-1)[\frac{\partial s_1}{\partial R_1} + M\frac{\partial s_1}{\partial R_2}].$$

Since $R_1 \leq R_2 \leq R_3$, f is strictly concave, $\frac{\partial s_i}{\partial y} < 0$ and $\frac{\partial^2 s_i}{\partial y \partial x_i} < 0$,

$$\frac{\partial s_1}{\partial R_2} = \frac{\partial s(R_1, y)}{\partial y} f'(R_2) \geq \frac{\partial s(R_j, y)}{\partial y} f'(R_1) = \frac{\partial s_j}{\partial R_1} \text{ with } j = 2, 3.$$

Thus,

$$\frac{\partial s_1}{\partial l_{12}} \leq (-1)[\frac{\partial s_1}{\partial R_1} + \frac{\partial s_2}{\partial R_1} + \frac{\partial s_3}{\partial R_1}] = 0.$$

This implies $l_{12} = 0$. We also have $l_{13} = 0$. Otherwise, by Step 3, $R_2 = 0$, and then $R_1 = 0$. But, by assumption, at least two inputs are positive.

Step 5. In an equilibrium where $R_1 \leq R_2 \leq R_3$, with a strict inequality $R_1 = 0$. Suppose that $R_1 > 0$. By Steps 3 and 4, $R_1 = 1 - 2(l_{21} + l_{31})$, $R_2 = 1 - l_{21}$, $R_3 = 1 - l_{31}$. Let us see that this can not be an equilibrium because agent 3 will be better off by reducing the input of agent 1 to cero. Let $l'_1 = l_1$, $l'_2 = l_2$, $l'_3 = l_{31} + \frac{R_1}{2}$. For this new strategy, $R'_1 = 0$, $R'_2 = R_2$, and $R'_3 = R_3 - \frac{R_1}{2}$. The consumption of agent 3 for those strategies is given by the following expressions:

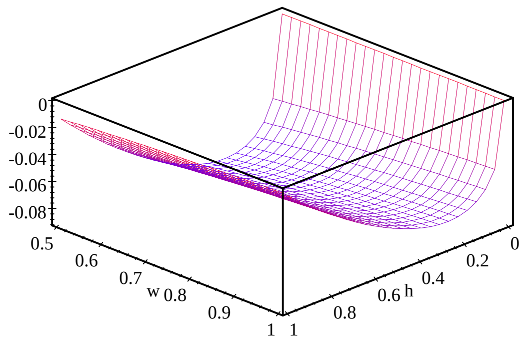
$$C_3(R_1, R_2, R_3) = \frac{(R_3)^{.5}}{(R_1)^{.5} + (R_2)^{.5} + (R_3)^{.5}}(R_1 + R_2 + R_3)^{.3},$$

$$C_3(R'_1, R'_2, R'_3) = \frac{(R_3 - \frac{R_1}{2})^{.5}}{(R_2)^{.5} + (R_3 - \frac{R_1}{2})^{.5}}(R_2 + R_3 - \frac{R_1}{2})^{.3}.$$

Let us define $w = \frac{R_2}{R_3}$, and $h = \frac{R_1}{R_3}$, since $R_1 \leq R_2 \leq R_3$, with a strict inequality and $R_1 > 0$, $0 < h < w$, and $\frac{1}{2} < w \leq 1$. Let

$$f(w, h) = \frac{1}{1 + w^{.5} + h^{.5}}(1 + w + h)^{.3} - \frac{(1 - \frac{h}{2})^{.5}}{(1 - \frac{h}{2})^{.5} + w^{.5}}(1 - \frac{h}{2} + w)^{.3}.$$

The next graph shows that this function is always negative for $0 < h < w$, and $\frac{1}{2} < w \leq 1$. Which implies that $C_3(R_1, R_2, R_3) < C_3(R_1', R_2', R_3')$.



$f(w, h)$

Step 6. There is no an equilibrium with $0 = R_1 < R_2 \leq R_3$.

Let us see that $0 < R_2 \leq R_3$ can not be an equilibrium because agent 3 can be better off by reducing the input of agent 2 to cero. The new vector of inputs is: $R_1' = 0$, $R_2' = 0$, and $R_3' = R_3 - \frac{R_2}{2}$. The consumption of agent 3 in both situations is given by:

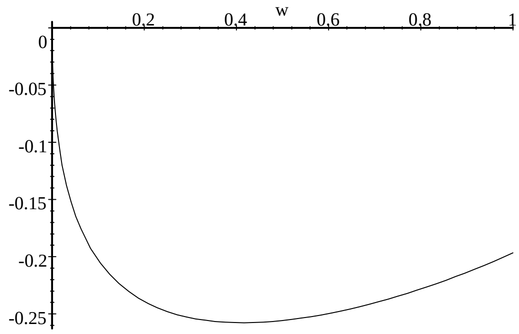
$$C_3(R_1, R_2, R_3) = \frac{(R_3)^{.5}}{(R_2)^{.5} + (R_3)^{.5}}(R_2 + R_3)^{.3},$$

$$C_3(R_1', R_2', R_3') = (R_3 - \frac{R_2}{2})^{.3}.$$

Let $w = \frac{R_2}{R_3}$, thus $0 \leq w \leq 1$. Let

$$g(w) = \frac{1}{w^{.5} + 1}(1 + w)^{.3} - (1 - \frac{w}{2})^{.3}.$$

The next graph shows that this function is always negative for $0 \leq w \leq 1$. Which implies that $C_3(R_1, R_2, R_3) < C_3(R_1', R_2', R_3')$.



$g(w)$

Therefore, there is no an equilibrium with at least two positive inputs. ■